Humanized yeast genetic interaction mapping predicts synthetic lethal interactions of FBXW7 in breast cancer

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Morgan William Bruce Kirzinger

©Morgan William Bruce Kirzinger, August 2018. All rights reserved.

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Head of the Department of Computer Science
University of Saskatchewan
176 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

A Synthetic lethal (SL) interaction involves a pair of genes (geneA and geneB) where inhibition of either geneA or geneB individually has no effect on cell viability, but the inhibition of both geneA and geneB causes cell death. SL interactions that occur between gene pairs can be exploited for cancer therapeutics. Studies in the model eukaryote yeast have identified approximately 550,000 negative genetic interactions that have been extensively applied to characterize novel pathways and gene functions. In the context of this thesis, a negative genetic interaction is the equivalent of a SL interaction. Harnessing the vast available knowledge of yeast genetics, we generated a Humanized Yeast Genetic Interaction Network (HYGIN) for 1,009 human genes with yeast orthologs and 10,419 interactions. Through the addition of patient-data from The Cancer Genome Atlas (TCGA), we generated a breast cancer specific subnetwork. Specifically, by comparing 1,009 genes in HYGIN to genes that were down-regulated in breast cancer, we identified 15 breast cancer genes with 130 potential SL interactions. Interestingly, 32 of the 130 predicted SL interactions occurred with *FBXW7*, a well-known tumor suppressor that functions as a substrate-recognition protein within the SKP/CUL1/F-Box ubiquitin ligase complex for degradation through the proteasome. Validation of these SL interactions using chemical genetic data indicate that patients with loss of *FBXW7* may respond to treatment with drugs like Selumitinib or Cabozantinib. Taken together, our patient-data driven interpretation of HYGIN represents a novel strategy to uncover therapeutically relevant drug targets.

Acknowledgements

I would like to thank Franco Vizeacoumar and Tony Kusalik for their never-ending feedback and support during my tenure at the University of Saskatchewan. They are the reason that I continue to contribute to bioinformatics research. I would also like to thank the Vizeacoumar research group and the Bioinformatics Lab at the University of Saskatchewan for their continual feedback and support throughout my Masters.

I would also like to thank J.B, S.C., M.D., T.K., and K.M. for their amazing support throughout my Masters, ensuring that I was well fed, making certain that I took time to relax and enjoy life, and most importantly encouraging me along the way. Finally, a huge shoutout to my parents and N.K.D., C.D., W.D., S.C. and D.K, R.H., K.N., B.P., K.S., and T.W. for keeping me sane, encouraging breaks for stress relief, and supporting this endeavour over the past two years.

to my parents

Contents

List of Tables

List of Figures

List of Abbreviations

APC/c	anaphase-promoting cyclosome complex
APC	adenomatous polyposis coli
ARF	alternate reading frame
ATM	ataxia-telangiectasia-mutated gene product
BL1	basal-like 1
BL2	basal-like 2
BLAST	basic local alignment search tool
BLIA	basal-like immune-activated
BLIS	basal-like immune-uppressed
BRCA	breast cancer protein
CD95	cluster of differentiation 95
CDC	cell division cycle
CCLE	cancer cell line encyclopedia
CPD	CDC4 phospho-degron motifs
CRISPR	clustered regularly interspaced short palindromic repeats
DAISY	data mining synthetic lethality
DECIPHER	database of genomic variation and phenotype in humans using ensembl resources
DNA	deoxyribonucleic acid
ER	estrogen rceptor
Fast-SL	fast synthetic lethal
FBA	flux balance analysis
FBXW7	F-box and WD repeat domain-containing 7
GO	gene ontology
HER2	human epidermal growth factor rceptor 2
HYGIN	humanized yeast genetic interaction network
IM	immunomodulatory
LAR	luminal androgen receptor
M	mesenchymal
MiSL	mining synthetic lethals
MSL	stem-like
PESCADOR	platform for exploration of significant concepts associated to co-occurrence relationships
PPP	pentose phosphate pathway
PR	progesterone rceptor
RNA	ribonucleic acid
RNA-seq	RNA-sequencing
RPE	ribulose-5-phosphate-3-epimerase
RSEM	RNA-seq by expectation maximization
SCNA	single copy number alterations

SCF	SKP/CUL1/F-Box) E3 ubiquitin ligase complex
SDL	synthetic dosage lethality
shRNA	short hairpin RNA
siRNA	small interfering RNA
SKP1	S-phase kinase associated protein 1
SL	synthetic lethality
SoF	survival of the fittest
SynLethDB	synthetic lethal database
T-ALL	T-cell acute lymphoblastic leukaemia
TCGA	the cancer genome atlas
TNBC	triple-negative breast cancer
TSG	tumour suppressing genes
UND	unstable
UniProt	universal protein reference

Chapter 1

Introduction

Genetically, cancer is a complex disease with no two patients exhibiting the same genetic profiles of tumors. Recent advances in tumor sequencing and identification of key driver genes has allowed the development of more targeted treatment strategies by leveraging individual patient genetics [10]. However, the druggability of these targets becomes challenging if these genes are not expressed or are down-regulated in certain cancers.

Synthetic lethality takes advantage of functional genetic interactions between gene pairs to develop targeted therapies. Genetic interactions are the phenomenon where two or more gene products interact with one another that results in a deviation of the expected phenotype [78]. Synthetic lethal (SL) interactions are beginning to be appreciated as a method for identifying druggable targets [30, 89, 101]. Inhibiting the synthetic lethal partner with a therapeutic drug selectively eliminates cancer cells leaving normal cells largely unaffected. By exploiting these synthetic lethal interactions, we can maximize the efficiency of personalized treatment options (novel drugs, or repurposed drugs) and, potentially, minimize the side effects that a patient experiences as a result of the therapy. Although the advent of genome-wide shRNA and CRISPR-Cas9 screens have provided the specificity needed to perform the comprehensive epistasis mapping on any number of chosen gene pairs [48, 90, 127], experimentally testing all possible gene pairs across multiple cell types to identify SL is laborious and time-consuming.

Previously, yeast SL genetic interaction data has been used to identify SL gene interactions in humans [27, 76, 113]. These yeast-directed approaches have helped uncover several druggable interactions in human cells. For example, McManus *et al.* identified the first humanized synthetic lethal interactions of *RAD54B*-deficient human colorectal cancer cells by mapping the corresponding yeast synthetic lethal interactions to human cancer cells [76]. Thus, rather than evaluating all human SL interactions, evaluating some in a yeast context

provided valuable insights into SL interactions in cancer. This approach reduces the dimensionality of searching for SL interactions in humans by reducing the total set of potential genes that can interact with one another to only those that have yeast orthologs. As a result, since the yeast genome is substantially smaller than the human genome and only a fraction of yeast genes have human orthologs, the total sample space for potential SL interactions in humans is reduced.

Although using pre-existing experimental data from yeast is extremely useful, results generated from mapping yeast data to humans does not provide a comprehensive picture of the human SL landscape. It turns out that the greatest strength of using yeast data to identify human SL interactions (reducing the dimensionality of the problem) is also it's greatest weakness: eliminating the majority of human genes from the analysis. The human genome contains approximately 20,000 genes and the yeast genome approximately 6,000. Given the evolutionary distance between yeast and humans, it is not surprising that there are only around 2,000 yeast genes that have human orthologs. As a result, if we are at most examining these 2,000 human genes with yeast orthologs for SL interactions, we are only evaluating 1/10 of the human genome. Therefore, using model species such as yeast is extremely useful for advancing the discovery of novel SL interactions in humans; however, caution must be taken when interpreting the results as there could exist many SL interactions between human genes without yeast orthologs.

The Cell Map contains a global genetic interaction map for *Saccharomyces cerevisiae* [23]. Using more than 23 million experimentally generated double mutants, approximately 550,000 negative and 350,000 positive interactions were reported in The Cell Map. Previous work by Deshpande *et al.* made use of an older version of this data [22], and was motivation for this study. Having identified potential problems with the Deshpande *et al.* work, this thesis aimed to improve upon their strategy for generating a human network of potential SL interactions based on previously experimentally validated yeast interaction data. Harnessing the vast available knowledge of yeast genetics and using yeast-human ortholog mapping, we generated a humanized yeast genetic interaction network (HYGIN) that has the potential to identify novel cancer-specific synthetic lethal interactions and novel cancer treatment strategies. Since no cancer data has been used in the generation of HYGIN any

cancer specific data can be incorporated to generate a subset of SL interactions in humans that are cancer specific. Breast cancer is the third most common type of cancer in Canada accounting for more than 25% of all new cases of cancer in women and 13% of all cancers in Canada. This network is then integrated with patient data from The Cancer Genome Atlas (TCGA) (`https://cancergenome.nih.gov/`) to identify genes that are down-regulated in breast cancer to provide a breast cancer-focused version of HYGIN. The breast cancer specific subnetwork contains 130 SL interactions with 15 genes that were identified to be down-regulated in breast cancer and 115 other genes in HYGIN. Of these 130 SL interactions, we predict 32 novel synthetic lethal interactions of the tumor suppressor *FBXW7* that could be exploited for patient-specific cancer therapeutics using previously identified or novel drugs.

The remainder of this thesis is organized as follows. Background information on cancer, yeast as a model organism, synthetic lethality, computational approaches to synthetic lethality, data sources for computational studies and *FBXW7* is given in Chapter 2. Chapter 3 articulates the research hypothesis of this thesis. The data used and methodology for generating HYGIN, initial data analysis techniques, statistical analyses used, validation of the network, and drug analyses are covered in Chapter 4. Results from evaluating HYGIN and the breast cancer subnetwork are presented in Chapter 5. Chapter 6 analyzes and interprets the results present in Chapter 5, discusses the shortcomings of previous work and compares HYGIN to it. Finally, Chapter 7 summarizes the conclusions of this work and proposes future research endeavours using HYGIN. Currently a manuscript describing this work is in the final stages of preparation for journal publication. The majority of the manuscript is integrated into this thesis document as Chapter 1, Sections 4.1 – 4.3, Sections 4.5 – 4.6, Chapter 5, portions of Section 6.1 and Section 6.4.

Chapter 2

Background

This section provides the necessary background material for understanding the research that is present in this thesis. Section 2.1 provides a brief introduction to cancer. Section 2.1.1 gives some background on the molecular classification of cancers, and section 2.1.2 provides an overview of information pertaining to breast cancer. Section 2.2 gives an introduction to yeast as a model organism, and section 2.2.1 discusses orthology and different ways that genes can be mapped between species. Section 2.3 contains a short background on synthetic lethality, it's applications and previous work. A survey of previous computational approaches to synthetic lethality can be found in section 2.4. Section 2.5 provides background information on the sources of data used in this thesis: 2.5.1; the Cell Map, 2.5.2; InParanoid, 2.5.3; SynLethDB, 2.5.4; the cancer genome atlas. Finally, section 2.6 provides an introduction to the tumor suppressor FBXW7 and it's role in cancer.

2.1 Cancer

Cancer is an extremely complex group of diseases that are the result of loss of control over cell division and cell growth. Recent statistics from the Canadian Cancer Society suggest that 1 in 2 Canadians will develop cancer in their lifetime, and 1 in 4 will die from it [107]. With such a high incidence rate in the population, it is necessary to identify and understand the mechanisms of cancer so that patients can receive better treatment. Since there are many different ways that a healthy cell can transform into a tumor cell, the identification of drug targets and personalized therapeutics is difficult. As a result, researchers are leveraging genetics research to develop personalized approaches to cancer therapy.

There are many different ways that cancer can arise in a cell leading to the complicated

task of identifying which genes are associated with which cancers. With advances in molecular biology techniques and genetics, researchers have identified two major types of genes that contribute to cancer: tumor suppressor genes (TSGs) and oncogenes. TSGs are a class of genes that under normal cellular function can repair damaged DNA (e.g. ataxia-telangiectasia-mutated gene product (*ATM*), breast cancer protein (*BRCA*), and *p53*), regulate cell division (e.g. alternate reading frame (*ARF*), *RIZ1*, *p27*, and *p53*), and control apoptosis (regulated cell death) (e.g. adenomatous polyposis coli (APC), cluster of differentiation 95 (*CD95*), and *p53*) [115]. However, in cases where TSGs are mutated or deleted these cell processes are no longer regulated and can lead to cancer. A classic example of a TSG is the nuclear transcription factor *p53* [87]. Genetic analysis of multiple cancer types suggests that *p53* has a loss of function mutation in more than 50% of cancers [5, 54, 82, 116]. In these cases, p53 loses the ability to repair damaged DNA and initiate cell apoptosis. There have even been studies in mice that show that mice with mutations to *p53* spontaneously develop tumors [31]. There are several different genetic means by which genes can lose expression and lead to cancer: mutations, deletions in the upstream region responsible for the regulation of a gene, whole gene deletions due to deletions on a chromosomal level, siRNA's, and most prominently DNA methylation resulting in a loss of gene expression [9, 33, 60, 79, 132].

On the other hand, oncogenes are the key drivers of tumor growth and are often expressed at high levels in tumor cells [67]. Oncogenes can cause chromosome instability and unscheduled cell growth and have been shown not only to be involved in cancer development, but also in cancer maintenance [125]. They have also been shown to play an important role in cancer cell metabolism where they aid tumors in meeting their nutritional retirements for growth, contributing to the notion that cancer is a metabolic disease [80]. However, oncogenes have to be activated in order to cause cancer. Proto-oncogenes are normally expressed genes in the cell that have the capacity to cause cancer when they are activated to become oncogenes [121]. For example, under normal cellular conditions, the highly regulated *MYC* (a transcription factor) proto-oncogene does not have the capacity to initiate cancer development. However, a mutation within the regulatory region for *MYC* could increase its expression making it an oncogene [25].

2.1.1 Molecular classification of cancers

Cancers tend to be classified based on the tissue or cell type that they were isolated from (e.g. breast, prostate, lung, esophageal, or pancreatic cancers). Though this method has been useful for doctors, patients, and researchers in the past, recent advances in genome sequencing technology have shed light onto the genetics of these cancers and potentially provided a complimentary means of classifying them for treatment [24, 53]. A recent study has shown that 43% of cancer cases on average are more similar to cancers from other tissue types than they are similar to other cancers from the same tissue [24]. For example, Heim *et. al* report that 14% of breast cancer cases are more closely related to ovarian cancer than they are to breast cancer. Furthermore, another study by Hoadley *et al* examined 12 cancer types and identified 11 major classification subtypes based on genetic analysis [53]. Although 5 of the subtypes contained cancer cases that were almost exactly the same as classifying cancers by tissue, it was found that cases of bladder cancer could be split into three different subtypes [53]. Classifying cancers into types based on genetics rather than tissue type provides a novel categorization for researchers that hopefully results in the re-purposing of cancer therapeutics from tissue types of cancer to molecular types.

Hoadley *et al.* released a more recent article in 2018 that expanded on their previous research and included 33 types of cancer and more than 10,000 tumors [52]. The updated research concluded that molecular tumor classification is influenced by cell-of-origin, but is not determined by it [52]. It is important to maintain the current cell-of-origin classification system for cancer as it allows researchers and doctors to understand what is happening in the body and how it is affecting physiology in a patient. However, molecular classification systems are also important as they provide doctors with novel treatment strategies that might not have been identified through cell-of-origin typing. Ultimately, neither of these two classifications systems is perfect; both have their merits and are necessary for prioritizing patient care and effective treatment.

Although cancer therapies have come a long way since they were first introduced, there are still negative side effects that impact patients undergoing treatment. As a result, there is a need for new therapeutics that specifically target cancer cells, while minimizing negative

side effects. By grouping cancers based on their genetic similarities, researchers can use existing therapeutic strategies that target a well know cancer-associated gene in one cancer and apply it to genetically similar cancers. Research efforts towards patient-specific therapies often evaluate known therapeutics and their effectiveness for patients with specific genetic profiles or biomarkers. This is an enticing research avenue that can potentially reduce the negative side effects of treatment and eliminates superfluous, ineffective treatments.

With advances in genome sequencing technology and the ability to identify cancer biomarkers through genome sequencing, personalized medicine is now becoming a reality. Patient-specific therapeutics are starting to become common practice in many cancer treatment strategies [124, 131]. These approaches involve selecting therapeutics based on the genetic profiles of the patients and their tumor(s). Identifying patient-specific novel druggable targets in cancer is still costly with respect to time and resources. As a result, research efforts are turning to computational techniques to predict patient specific druggable genes for efficient cancer therapeutics.

2.1.2 Breast cancer

Breast cancer is the third most common form of cancer in Canada, accounting for more than 25% of new cases of cancer in women in 2017 [107]. There are 5 intrinsic types of breast cancer: luminal A, luminal B, triple-negative breast cancer (TNBC), human epidermal growth factor receptor 2 enriched (HER2-enriched), and normal-like [14, 85, 93, 110]. Luminal A breast cancer is characterized by the presence of the estrogen receptor (ER) and the progesterone receptor (PR), the absence of HER2 (abbreviated ER+/PR+/HER2-), and low levels of the Ki-67 protein. This type of breast cancer has the best prognosis among breast cancers [85]. Luminal B breast cancer is ER+/PR+, either HER2 positive or negative (ER+/PR+/HER2+/-) and has high levels of Ki-67 [85]. In contrast, TNCB is ER-/PR-/HER2- and represents 10-20% of breast cancers [135]. HER2-enriched breast cancer is ER-/PR-/HER2+ and tends to grow faster and have a worse prognosis than luminal A and luminal B breast cancers [85]. Finally, normal-like breast cancer is ER+/PR+/HER2- and very similar to luminal A except it has a slightly worse prognosis than luminal A breast cancer [85]. Within each type of breast cancer mentioned above, there are various sub-types defined by their molecular profiles. Of

particular interest due to its high prevalence among women in Canada is TNBC.

Over the past decade, researchers have been working to identify the molecular subtypes of TNBC. In 2011, Lehmann *et al.* published a study that identified six molecular subtypes of TNCB: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR) [68]. However, in 2016, Lehmann updated their classification system because transcripts previously described in IM samples were from infiltrating lymphocytes, and transcripts from MSL samples were actually from tumor-associated stromal cells [69]. As as a result, the new classification system that they propose contains only four of the above mentioned molecular subtypes of TNCB: BL1, BL2, M, and LAR [69].

Interestingly, between the publishing of the Lehmann 2011 [68] and Lehmann 2016 [69] manuscripts, three other research groups published different molecular classification systems for TNBC [16, 59, 73]. Masuda *et al.* reported seven subtypes in 2013: BL1, BL2, M, IM, MSL, LAR, and unstable (UNS) [73]. Burstein *et al.* 2015 reported four molecular subtypes which they named LAR, mesenchymal (MES), basal-like immune-uppressed (BLIS), and basal-like immune-activated (BLIA) [16]. In contrast, Jezequel *et al.* identified three molecular subtypes: LAR, and two different types of basal-like (one with high M2-like macrophages and a low immune response, and the other with low M2-like macrophages and a high immune response [59]).

There is an unsurmountable amount of evidence that strongly suggests that TNBC is heterogeneous and made up of several different subtypes. Although there does not seem to be a consensus on the molecular subtypes within the research community, it is reassuring that most studies are identifying some of the same subtypes: basal-like, MES, and LAR. Identifying molecular subtypes of breast cancer has broader applications for the cancer research community. By expanding molecular subtyping analyses to multiple cancer types, researchers can group cancers based on genetics rather than cell-of-origin. The hope is that this novel method of classification will group cancers into types that are treatable by repurposing existing therapeutics and increase patient care.

2.1.3 The tumor suppressor FBXW7

FBXW7 (F-box and WD repeat domain-containing 7) is the substrate recognition component of the SCF (SKP/CUL1/F-Box) E3 ubiquitin ligase complex that has several aliases: *FBW7*, *AGO*, *CDC4*, and *SEL10* (Figure 2.1) [130]. FBXW7 has been classified as a tumor suppressor [130], and studies have shown that loss of *FBXW7* gene expression results in chromosomal instability through deregulation of its protein substrates, including cyclin E [111]. When conserved CDC4 phospho-degron motifs (CPDs) on target proteins are phosphorylated, FBXW7 binds to them and the protein substrate is targeted for ubituitination through the SCF complex. Because the substrates of FBXW7 are involved with cell division, cell growth, and cell differentiation, phosphorylation of these substrates by glycogen synthase kinase 3 (GSK3) is highly regulated [26]. The proteins that FBXW7 interacts with include several well known oncogenes including cyclin E, JUN, Notch, and MYC [130].

The *FBXW7* gene is located on human chromosome 4q32, a location that is often deleted in cancers [112]. Because FBXW7 is part of the SCF E3 ubiquitin ligase complex, in cancers where it is deleted, its oncogenic substrates become excess in the cell. For this reason, FBXW7 was proposed to function as a tumor suppressor [129]. It has also been shown that *FBXW7* is mutated in around 6% of cancer tumors but with a large amount of variation between tumor types. For example, the cancer types most frequently mutated are T-cell acute lymphoblastic leukaemia (T-ALL) and cholangiocarcinoma with 31% and 35% of cases harbouring a mutation in *FBXW7* respectively [1]. In contrast, prostate cancers has a 6% mutation rate [63].

The FBXW7 protein contains an F-box protein interaction domain that interacts directly with SKP1 (S-phase kinase associated protein 1) and is responsible for the recruitment of ubiquitination substrates to the SCF [62]. It also contains eight WD40 repeats that are used to interact with its phosphorylated protein substrates at the CPD domain [46], and a D domain that is used for the dimerization of FBXW7 [129]. The third and fourth WD repeat domains of FBXW7 contain 3 conserved arginine residues that interact with the phosphorylated substrates, while the remaining repeats are responsible for substrate binding [46, 86].

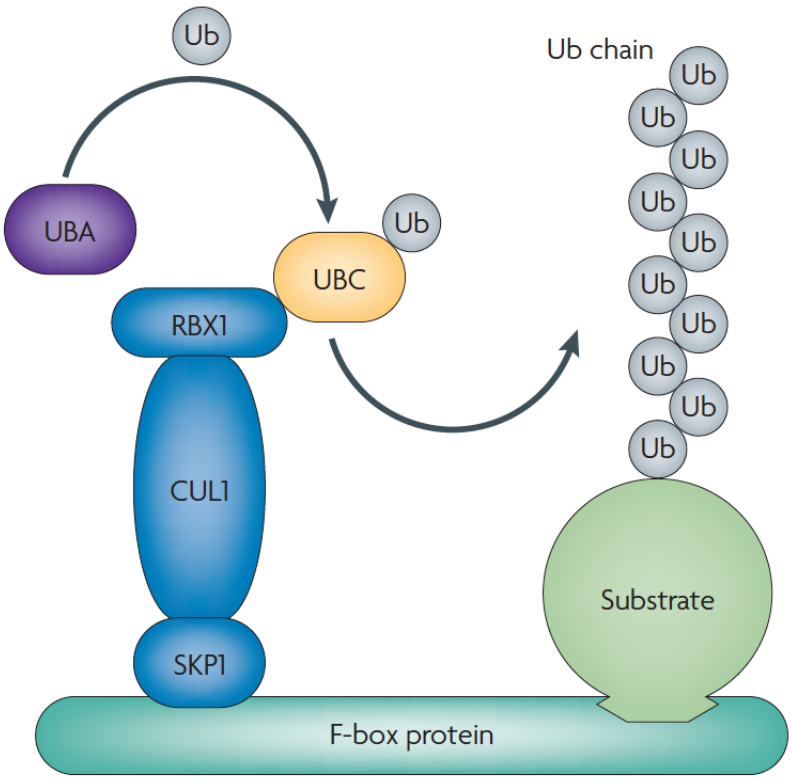


Figure 1 | **Schematic of an E3 SCF (complex of SKP1, CUL1 and F-box protein) ubiquitin ligase.** The SKP1 (S-phase kinase-associated protein 1)–CUL1 (cullin 1)–RBX1 (RING box 1) core complex recruits substrates through interchangeable substrate-specific F-box proteins. RBX1 binds to the E2 ubiquitin-conjugating enzyme (UBC) that was previously charged with ubiquitin (Ub) by an E1 ubiquitin-activating enzyme (UBA). Polyubiquitylated substrates are then targeted for destruction by the proteasome.

Figure 2.1: A schematic diagram of the SCF protein complex reprinted with permission from Nature Reviews Genetics (Appendix C.1) [130].

Given that FBXW7 is responsible for the degradation of several oncogenes, it is not surprising that targeting FBXW7 and it's substrates could be a viable approach to treating cancers. Targeting FBXW7 directly is one option given that a large portion of tumors maintain *FBXW7* expression. However, in cancers where *FBXW7* is lost, a better approach might be to target the oncogenic substrates of FBXW7. This has already been achieved for two substrates of FBXW7: in human T-ALL Notch has been targeted using γ-secretase inhibitors, and in murine sarcomas, MYC has been targeted using γ-secretase inhibitors as well [42, 57]. Another potential avenue for targeting cancers through loss of *FBXW7* is to evaluate SL interactions with *FBXW7* and use those genes as potential drug targets. This strategy would be particularly useful for cancers with low expression of *FBXW7* since it's SL partners can be targeted with drugs on a patient by patient basis.

2.2 Yeast as a model organism

Model species are often extremely useful starting points for research. In the case of yeast, there are several studies that support its utility and application to human research, especially cancer research [11, 43, 75, 91, 92, 95, 97, 104]. Hartwell was the first to discover genes involved in the yeast cell cycle [49, 50, 95]. These cell division cycle (CDC) genes in yeast have been studied extensively in a human cancer context [8, 17, 71, 99], and mutations in these genes in humans have been shown to be involved in cancer [95]. In addition to having genes that are relevant to studying human cancer, yeast is also easily cultured in a laboratory and has a relatively small genome [41]. Furthermore, there are currently many research groups focusing on yeast resulting in large amounts of data that can be mined and applied to human cancer research [11, 43, 75, 91, 92, 95, 97, 104].

2.2.1 Ortholog mapping

Prior to discussing the ways of mapping yeast genes to human genes, it is important to identify and define the following key evolutionary terms: homologs, orthologs, and paralogs. Homologs are genes with a common evolutionary origin but not necessarily the same function and can be divided into orthologs and paralogs. Orthologs are genes with a common ancestor

that are separated by speciation and have kept the same function over time. In contrast, paralogs are genes with different functions in the same genome related by duplication. There are two main ways that ortholog mapping is conducted *in silico*: sequence similarity (for example using a BLAST-based approach (basic local alignment search tool) [56, 70, 109]) and a phylogeny-based approach [55, 96, 100, 141]. The BLAST-based approach involves pairwise sequence comparison and is a fast approach for identifying potential orthologs. In contrast, the phylogeny-based approaches compare gene trees to species trees to identify gene duplication events and predict orthoogy. Furthermore, phylogeny-based approaches are more time- and resource-intensive but are useful in identifying more fine-grained ortholog relationships. InParanoid was built using a sequence similarity approach and has an easy-to-use interface.

2.3 Synthetic lethality

There are many different genetic interactions that researchers can exploit in order to identify potential genes that would act as drug targets for cancer therapeutics. Genetic interactions are the phenomenon where two or more gene products interact with one another that results in a deviation of the expected phenotype [78]. Two such genetic interactions are synthetic lethality and synthetic dosage lethality. A SL interaction involves a pair of genes (geneA and geneB) where inhibition of either geneA or geneB individually has no effect on cell viability, but the inhibition of both geneA and geneB causes cell death. In contrast, a synthetic dosage lethal (SDL) interaction involves a pair of genes (geneA and geneB) where over-expression of either geneA or geneB individually has no effect on cell viability, but the inhibition of geneA when geneB is over-expressed causes cell death (Figure 2.1). The concept of synthetic lethality was first described in 1922 by Calvin Bridges [15]; however, recent applications to cancer and the discovery of complex SL interactions (such as SL triples and quadruples [74]: SL interactions where all three or four genes need to be down-regulated for the lethal phenotype) are more recent applications of the concept. There are three proposed mechanisms of SL interactions: (1) each gene is involved in parallel redundant cellular pathways, (2) genes are members of distinct cellular pathways that generate the same product, and (3) the genes are

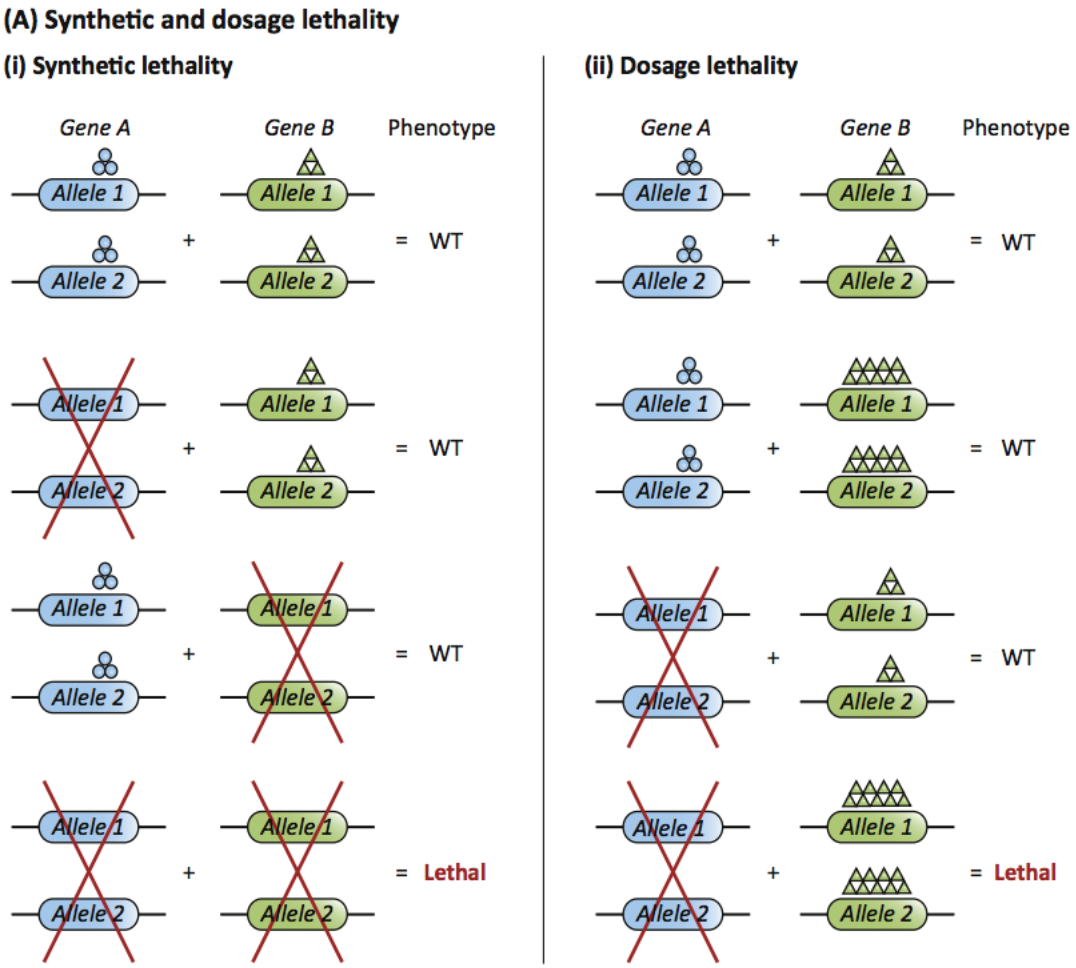


Figure 2.2: A schematic diagram of SL and SDL reprinted with permission from Cell Press (Appendix C.2) [89]. (i) represents a schematic of SL and (ii) is a schematic representation of SDL.

translated into different subunits of the same protein complex (Figure 2.2) [34].

Synthetic lethality has a wide array of applications in microbial, plant, animal, and human research. The most effective means of accurately identifying SL interactions is through *in vitro* assays that knock down both genes and monitor cell survival. This is often done *in vitro* and *in vivo* using CRISPR-Cas9 [102] or RNA interference [32]. Despite similar methods for identifying SL interactions, applications in microbes, plants, and animals (including *Drosophila melanogaster* and *Caenorhabditis elegans*) are different from cancer applications in humans. For example, in bacteria SL interactions have been evaluated to determine essential plasticity (the SL interaction works as a backup mechanism, e.g. Cell Envelope

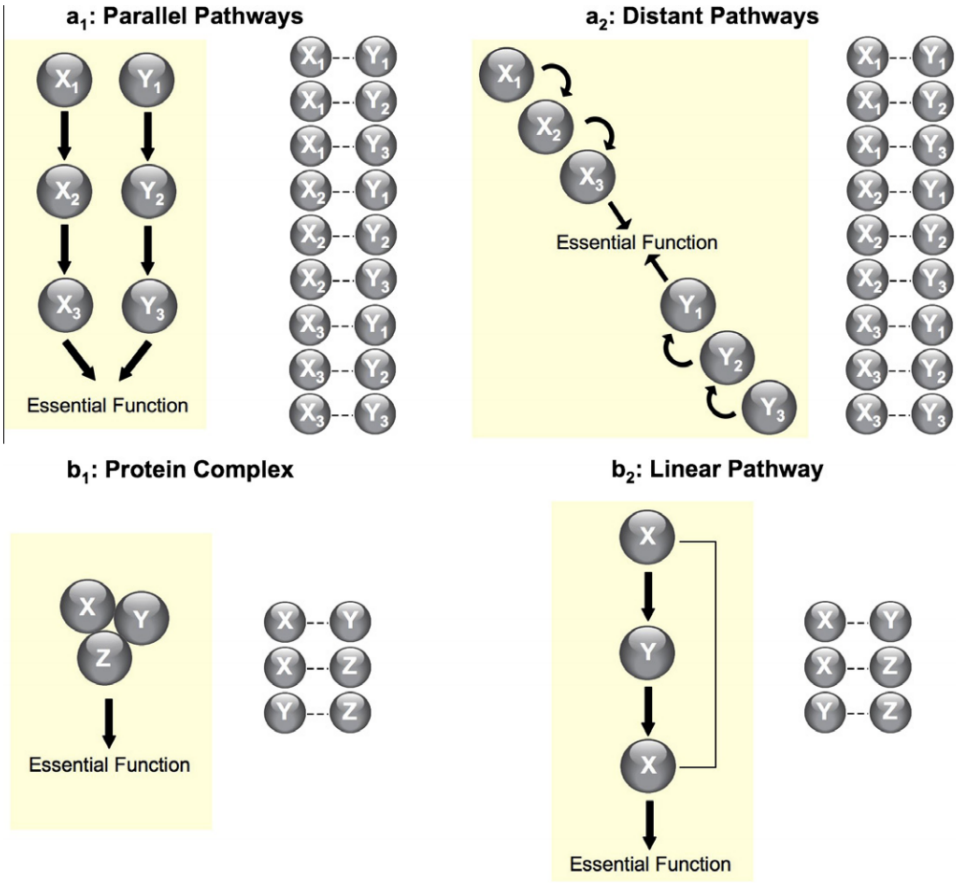


Fig. 1 – Different mechanisms leading to synthetic lethality. The possible SL gene pairs are reported on the right side of each panel. Synthetic lethality can arise form the absence of two genes acting in redundant parallel pathways (a_1) or distant pathways (a_2). Alternatively, it can originate from the lack of two subunits of an essential protein complex (b_1) or two proteins of the same essential pathway (b_2).

Figure 2.3: A schematic diagram of mechanisms of SL reprinted with permission from European Journal of Cancer (Appendix C.3) [34].

Biosynthesis as an essential backup for Membrane Lipid Metabolism) and essential redundancy (the SL interaction works as a parallel mechanism - shown by SL interactions affect a single function or pathway e.g. apoptosis) in metabolic networks [44]. In other microbial studies, SL interactions are used to identify gene functions in metabolic pathways [117], predict flux distribution and metabolic function for increased production of biofuels [51], analyze metabolites in known pathogens [40], and discover antibiotic combinations [3]. Furthermore, a large number of studies have focused on SL interactions in yeast, specifically *Saccharomyces cerevisiae*. These studies include mining protein-protein interaction networks for SL interactions [88], screening for genome instability [139], flux balance analysis (a mathematical model for simulating metabolism) to identify SL interactions relevant to metabolic pathways [47], identifying genes required for chromosomal segregation [77], identifying complex molecular interactions through SL interactions in multi-protein complexes [65], gene function prediction [138], and examining the total set of SL interactions in the yeast genome [22, 23, 84].

Although *in vitro* assays are important for validating SL interactions, they can be time-consuming, expensive, and labour-intensive, especially if the genome of the organism being studied is large in size. For example, studying SL interactions *in vitro* in human cancers requires extensive cell line work and the generation of millions of double-mutant cell lines in order to evaluate all possible SL interactions. Given approximately 20,000 genes in the human genome, there are almost 200,000,000 genetic interactions in an exhaustive *in vitro* screen. To that effect, identifying higher order SL triples or quadruples *in vitro* becomes increasingly more expensive and time-consuming as the dimensionality of the problem increases [74]. As a result, researchers are beginning to exploit SL interactions in model organisms (such as yeast [22, 23, 84]) to identify promising SL candidate drug targets. These approaches are more time- and cost-effective since they reduce the number of potential SL interactions to a feasible set that can be evaluated *in vitro* in a reasonable amount of time. Additionally, computational studies are working on developing predictive models for synthetic lethality that will serve as a form of dimensional reduction, which means that *in vitro* resources can focus on validating SL interactions that have already been computationally predicted [27, 58, 74, 105, 114]. Therefore, it is not surprising that there is an increased interest in computational approaches to identifying SL interactions in humans that make use of previously identified SL interactions

in model species.

2.4 Computational approaches for identifying SL interactions

Often, if geneA is mutated, down-regulated, or not expressed in cancer and it is in a SL interaction with geneB, then geneB becomes a potential target for novel or existing drugs. This type of targeted drug-gene interaction can reduce the side effects of the drug, increase the efficiency of treatment, and increase the comfort of the patient during treatment. It can also prevent patients from being treated with drugs that will not be as effective for their cancer, reducing the cost and length of patient treatment and potentially preventing negative side effects like liver damage from excessive (or unnecessary) chemotherapy. Computational approaches to synthetic lethality in cancer are increasing in popularity using existing cell line gene expression data, patient survival and mutation data, and previously experimentally validated SL interaction data from model organisms (e.g. yeast) [27, 58, 74, 94, 114].

Some programs are designed just for the purpose of identifying SL interactions in microbes, such as Fast-SL [94]. Fast-SL takes a linear programming approach to flux balance analysis (FBA). FBA uses metabolic models to analyze fluxes in a network and analyzes gene enrichment analysis, allowing for gene essentiality analysis and potential identification of SL interactions [120]. Fast-SL was used to analyze *Escherichia coli*, *Salmonella enterica* Typhimurium and *Mycobacterium tuberculosis* for SL interactions. It not only identified SL gene deletions in all three species but also identified novel SL triples. This approach to SL identification has applications in the identification of novel antibiotic targets and can help researchers explore complex genetic interactions in pathogens.

In addition to studies that are focused solely on bacteria and yeast, there are also several studies that identify SL interactions in yeast with the intention of applying them to human cancer, or use existing experimentally validated yeast SL interactions to identify cancer relevant SL interactions. Using yeast and other species as model organisms for studying SL interactions in humans can be extremely valuable, resource efficient, and help reduce the sample size for *in vitro* and *in vivo* testing. For example, Meta-SL is a computational

approach to identify SL gene pairs in yeast combining information from 17 genomic and proteomic features with the results of several classification models [133]. Upon the identification of SL interactions in yeast, ortholog mapping from yeast to humans revealed a set of cancer relevant SL interactions to study further.

Other computational studies use previously identified yeast SL interactions to identify potential SL interactions in human cancer [27]. Deshpande *et al.* use existing yeast SL interaction data from The Cell Map [22], ortholog mapping from yeast to humans, and cancer mutation data in order to identify potential cancer relevant SL interactions. This study identified 1,522 SL interactions relevant to cancer and tested SL interactions between *SMARCB1* and *PSMA4*, and *ASPSCR1* and *PSMC2* as proof of concept. More than 500 SL interactions containing at least one gene that is known to be associated with cancer were identified in this study.

In contrast, Srivas *et al.* identified a list of human cancer genes, mapped them to yeast orthologs and their SL interactions, and tested the resulting SL interactions in human cell lines [114]. This strategy is arguably more cancer directed as it starts with a list of known human tumor suppressor genes (TSG) and applies existing yeast genetic interaction data. Though the results are highly relevant to human cancer, they are extremely focused on known TSGs and there is the potential for the exclusion of SL interactions that might involve a previously unidentified human TSG. Despite their differences, both of these research endeavours focus on previously identified genes associated with cancer.

In contrast, the identification of SL interactions in humans without the use of a model organism is complicated by the size of the genome, which greatly increases the number of possible SL interactions. As a result, a number of specialized programs have been implemented for the identification of SL interactions specifically in human cancer. One of the more recent research efforts by Matlak and Szczurek [74] introduces SurvLRT, a program that analyzes cancer patient data using a statistical likelihood ratio test in order to identify SL gene pairs and triples in humans. SurvLRT is the first computational approach that identifies biomarkers for SL drug therapy, and was successful in identifying *TP53BP1* as a biomarker for the well established *PARP-BRCA1* SL gene pair. What's more, SurvLRT identified a novel biomarker for the same SL interaction that has a higher number of deletions

in cancer patients than *TP53BP1*, opening new research avenues for personalized medicine in cancer therapy.

Another computational pipeline for the identification of SL and SDL interactions in human cancer is DAISY (data mining synthetic lethality identification pipeline) [58]. DAISY makes use of multiple cancer patient and cancer cell line datasets and uses three statistical inference strategies to predict SL interactions in cancer. The first inference strategy, genomic survival of the fittest (SoF), analyzes somatic copy number alterations (SCNA) and mutation profiles for gene pairs where gene B has higher SCNA values and where geneA is inactive. The next strategy, shRNA-based functional examination, uses gene essentiality profiles to determine if, for each gene pair, geneB has lower shRNA expression in samples where geneA is inactive. Finally, DAISY evaluates pairwise gene co-expression using gene expression profiles to identify gene pairs whose expression is positively correlated. Gene pairs that are statistically significant in all three of the inference strategies are reported as positive SL interactions in human cancer.

Unfortunately, because DAISY uses a small number of mutations and shRNA data from existing cell lines, it misses all SL interactions that are false negatives in shRNA screens due to a lack of representation in existing cancer cell lines or due to incomplete knockdown. To counter these shortcomings, Sinha *et al.* [105] developed MiSL (mining synthetic lethals) to identify cancer specific SL interactions in specific cancer types. The MiSL algorithm is based on Boolean implications mined from pan-cancer datasets. Because MiSL takes into account all types of mutations, compared to DAISY which only takes into account frame shift and non-sense mutations, MiSL out performs DAISY by identifying several novel SL interactions that DAISY can't identify. This does not mean that MiSL is perfect; rather, it shows that research is constantly being completed to improve computational approaches for SL identification in humans.

Since these computational methods rely on the use of existing datasets, there is still a need to validate their conclusions *in vitro* (and sometimes *in vivo*) to ensure that the computational method is accurate. However, with an increasing number of SL interactions identified through cell line testing and a database for synthetic lethal interactions (SynLethDB [45]), it is now possible to validate *in silico* approaches by comparing the results to SL interactions that

have been previously identified in the lab and by other SL computational approaches. This may pose challenges though as a novel computational method of identifying SL interactions in humans might have no SL interactions in common with previous research and the predicted SL interactions have to be validated through *in vitro* and *in vivo* studies. However, results that have overlap with existing computational methods can be partially validated by analyzing those interactions. *In vitro* and *in vivo* studies can then provide additional support for the interactions that overlap with existing studies.

Unfortunately, there only a few databases that contain human SL interaction data from multiple studies. The most well known is the SynLethDB [45], which contains SL interactions from large scale computational studies, SL interactions recovered from text mining, and computationally derived SL interactions from the DECIPHER database [36]. Unfortunately, there are errors in this database (see Results and Discussion), which result in a loss of confidence in the text mining entries and reduction in the number of SL interactions available. However, SynLethDB is a starting point for comparing known SL interactions to those identified in novel computational studies.

This thesis combines previously identified yeast genetic interaction data with yeast-human orthology data to develop a novel network of human SL interactions that can be computationally verified using existing studies. As mentioned earlier, Deshpande *et al.* [27] used existing SL interaction data from yeast to identify SL interactions in human cancer (this topic expanded on in section 6.1). The research described here is based on their approach; however, updated databases and patient gene expression data from The Cancer Genome Atlas (TCGA) is used to generate a breast cancer specific set of SL interactions in humans. Using this information in conjunction with SynLethDB entries, we show that the network of predicted SL interactions in humans is not the result of random chance and further identify 32 druggable gene targets specific to breast cancer.

2.5 Data sources for computational studies

In order to use a computational approach to identify human SL interactions previous data and experimental results were necessary to begin the analyses. The main sources of data for

the work presented in this thesis were: The Cell Map [123], the TCGA database (`https://cancergenome.nih.gov/`), SynLethDB [45], and the InParanoid database [109]. In addition to these main data sources several additional data sources were used such as: cancerRXgene database [136], the cancer cell line encyclopedia (CCLE) [7], the Ensembl database [141], and the work of Deshpande *et al.* [27] in order to curate, compare and analyze the data in this thesis.

2.5.1 The Cell Map

As previously discussed, yeast is a good model organism for studying cancer in humans (Section 2.2). The starting dataset for the generation of the HYGIN network was The Cell Map [123]. Cell Map is a current dataset of previously biologically validated yeast genetic interactions that was last updated May 2016. The dataset covers close to 90% of yeast genes and contains approximately 550,000 negative and 350,000 positive yeast genetic interactions. In the Cell Map dataset, deletion query mutant strains were screened against deletion mutant arrays in the search for genetic interactions. Interactions were scored based on an ϵ value (a score for genetic interactions[1]) and statistical significance. Using the data cut-off values for P-value and ϵ as described in the work by Deshpande *et al.* [27], a subset of the Cell Map data was used to generate the HYGIN network.

2.5.2 InParanoid

Orthologous mapping of yeast to human genes used the InParanoid database of orthologs [109]. As mentioned previously, the InParanoid database contains 273 species, 37,128 species pairs, and 2,999,062 orthologous proteins including 2,041 orthologs between yeast and humans [109]. InParanoid uses sequence similarity implemented through a BLAST-based approach to ortholog mapping and accounts for one-to-one and one-to-many orthologs between pairs of species. InParanoid uses bootstrap confidence values to assign a confidence score that the seed used for the comparison and it's ortholog are true orthologs. InParanoid version 8 is the

[1]The specific details for how ϵ was calculated were not provided in the original manuscript, nor were explicit details given in the original Science paper [122].

most current version [109], and it is updated regularly to include new genetic information for identifying orthologs.

2.5.3 The synthetic lethal database

SynLethDB is a curated database that contains SL gene pairs from other databases, biochemical results, text mining, and computational predictions [45]. In addition to 19,952 human SL interactions, the SynLethDB contains SL interactions for *Saccharomyces cerevisiae* (13,241), *Drosophila melanogaster* (423), *Mus musculus* (366), and *Caenorhabditis elegans* (107). Two of the major contributors to the human SL interactions in the SynLethDB are a computational study, DAISY (discussed in Section 2.3) [58], and SL interactions that were identified through an analysis of the DECIPHER database [36]. The remaining major contributors to the SynLethDB are other smaller *in vitro* analyses and SL interactions identified through a text-mining algorithm. The text mining approach that Gou *et al.* used for SynLethDB involved three major steps: search, train, and extract. During the search step, the titles of the articles in the PubMed database were searched for the terms "synthetic lethal" and "synthetic lethality". The abstracts of the 331 articles that were identified to contain these terms in their titles were used in to train the literature ranking tool, MedlineRanker [37]. MedlineRanker uses a defined training set (e.g. the previously mentioned 331 abstracts) to rank the literature in a database according to how relevant they are to synthetic lethality. Gou *et al.* chose to use the top 1,000 articles from the past 10 years in their search for SL interactions from previous literature. The third step in their text mining procedure was to use PESCADOR (Platform for Exploration of Significant Concepts AssociateD to co-Occurrence Relationships) [6], an online tool for identifying biointeractions from articles in the PubMed database. Based on the information from MedlineRanker, PESCADOR used keywords to identify SL interactions in the literature. Although the text mining portion of the SynLethDB contained several errors (see Section 5.2 for greater detail and examples), text mining approaches are a valuable means of populating databases and consolidating information for *in silico* studies provided that checks and balances are in place. The authors report that they will address the errors in the upcoming version of the database.

2.5.4 The cancer genome atlas

The cancer genome atlas is currently one of the most up-to-date comprehensive publicly available cancer genetics databases (`https://cancergenome.nih.gov/`). In addition to cancer gene expression data for more than 10,000 cancer samples across 33 cancer types, the TCGA also contains mutation data, patient data, and raw sequencing data. Additionally, the TCGA contains non-cancerous gene expression data for some cancer patients, including 114 normal breast tissue samples from 114 breast cancer patients. This paired data, i.e. gene expression data from normal tissue and cancerous tissue from the same patient, is extremely valuable for researchers. Specifically, it gives researchers the ability to evaluate differential gene expression in cancer versus normal tissue while accounting for baseline gene expression in individual patients.

Chapter 3

Research Hypothesis

Although the principle of using bioinformatics techniques for the identification of SL interactions *in silico* is not new, there are ways that we can improve upon previous techniques and approaches to better identify SL interactions in breast cancer. With a constant increase in the amount of patient genomic data and an ever-increasing number of cancer cell lines (and genomic data for these cell lines), novel practical approaches are needed for the analysis of this data. Furthermore, there are also a number of studies that have started with experimentally validated SL interactions in non-human model organisms and used this data as a spring board for the identification of SL interactions in human cancer. This thesis describes an improved approach using yeast orthologs to develop a humanized genetic interaction network in a cancer independent context.

This thesis used yeast genetic interaction data from The Cell Map [23] and cancer gene expression data from the TCGA to identify breast cancer specific SL interactions. We hypothesize that the creation of an improved and updated humanized yeast genetic interaction network based on yeast-human orthologs and the application of breast cancer gene expression data will allow us to identify genes for targeted breast cancer therapeutics. The pursuit of this hypothesis was broken down into the following goals:

1. Curate the existing yeast genetic interaction network from The Cell Map, and convert it to a humanized genetic interaction network that uses gene identifiers that can stand the test of time with minimal changes.

2. Identify the set of genes that are down-regulated in breast cancer patients using TCGA data.

3. Map the set of down-regulated genes onto the humanized yeast genetic interaction

network to generate a breast cancer specific subnetwork.

4. Identify genes that have SL interactions with multiple cancer related genes that are down-regulated in breast cancer.

5. Identify potential novel drugs for target genes using the drug-gene interaction database.

The goal of this project is to generate a humanized yeast genetic interaction network (HYGIN) based on human orthologs in yeast in a cancer independent context, apply cancer gene expression data, and identify novel SL gene pairs for targeted breast cancer therapeutics.

Chapter 4

Data and Methodology

Using yeast genetic interaction data as the basis for identifying SL interactions in humans is not a novel concept. Previously, the work of Deshpande *et al.* used yeast genetic interaction data to generate a cancer-independent human SL interaction network [27]. However, their network used all yeast-human orthologs including instances of one-to-many and many-to-many mapping. Using an updated version of the yeast interaction dataset [23], and working to improve on the strategy of Deshpande *et al.*, this thesis generates a stringent humanized yeast genetic interaction network that uses strict one-to-one yeast-human ortholog mapping. We then use breast cancer data from the TCGA and generate a breast cancer specific subnetwork that contains potential gene targets for breast cancer therapeutics.

4.1 Generating the humanized genetic interaction network

In order to generate HYGIN, experimentally validated yeast interactions from the May 2016 release of The Cell Map [23] were used. The yeast network contains quantitative genetic interactions for all gene pairs in *Saccharomyces cerevisiae* and is the result of double mutant arrays conducted in yeast that represent nearly all of the SL interactions in the yeast genome.

The yeast interaction data contains approximately 550,000 negative genetic interactions (SL interactions) between approximately 90% of the genes in the yeast genome [23]. In order to generate a humanized network, the yeast network was translated in three stages: yeast gene name to yeast protein, yeast protein to human orthologous protein, and finally human protein to human gene name. The UniProt database [2] was used for the first and last stages

and InParanoid Version 8.0 [109], an online tool for identifying orthologs between two species, was used to identify the human orthologs of the yeast proteins in stage two. Strict one-to-one mapping with InParanoid was used for the network to prevent ambiguities in the translation process (Supplementary Table B.1).

For example, The Cell Map contains the synthetic lethal interaction between the yeast genes *RPN1* and *MET30*. The names of these two genes were converted to yeast protein accession numbers (P38764 and P39014 respectively). InParanoid ortholog mapping then mapped the yeast protein accession numbers to human protein accession numbers (Q13200 and Q969H0 respectively). Finally, the human protein accession numbers were mapped to the corresponding human genes *PSMD2* and *FBXW7*, and a SL interaction between these genes was added to the HYGIN network. Supplementary Table B.2 lists tall the interactions in HYGIN.

Performing the mapping and filtering steps is tedious and error prone. Therefore the process of generating HYGIN was automated and performed computationally using a combination of bash (UNIX shell) and Python scripts. The network information was imported into Cytoscape [103] for visualization and analysis. The resulting network has 1,009 nodes (human genes) and 10,419 edges (proposed SL interactions between human genes). A summary of this process can be found in Figure 4.1.

4.2 Creating the breast cancer specific subnetwork

HYGIN was developed in a cancer independent manner in order to avoid certain biases such as starting with known oncogenes or TSGs, or limiting yeast ortholog mapping to only known cancer associated genes in humans. As a result, HYGIN is disease independent and the SL interactions in the network could be used for multiple different studies. However, our research hypothesis involved identifying SL interactions that are relevant to breast cancer. As a result, breast cancer data needed to be incorporated into HYGIN in order to identify potential gene targets for breast cancer therapeutics. This was accomplished through the integration of TCGA breast cancer data with HYGIN and the generation of a breast cancer specific subnetwork. The TCGA contains gene expression data for both normal (114 samples)

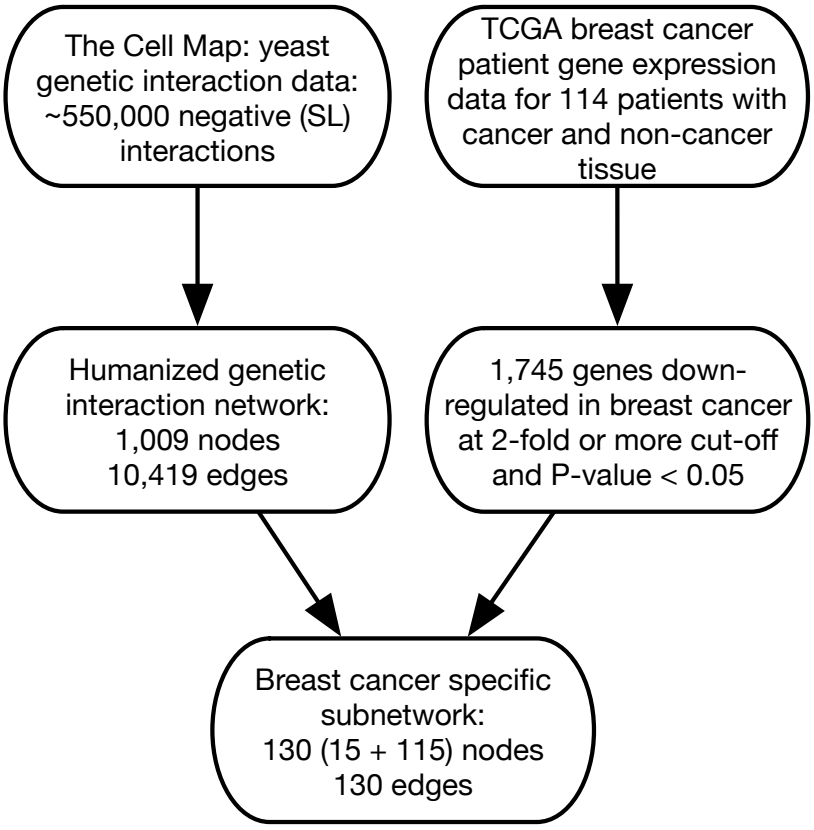


Figure 4.1: Flow diagram showing the flow of data and method used to generate the breast cancer specific subnetwork.

and cancerous (1,104 samples) tissue.

Some studies use all available data as a means of reaching statistically significant results, or as a means of completing an exhaustive analysis [58, 106], for example, using pooled normal tissue data and comparing it to pooled cancers tissue data as a means of statistically evaluating changes in gene expression. However, we used a more stringent methodology and only analyzed TCGA patient data for which there was gene expression data for both breast cancer tissue and normal breast tissue from the same patient. This increases the confidence in determining gene up- and down-regulation as it helps eliminate false positive results that occur due to differences in basal gene expression between individuals. By comparing up- and down-regulation of genes between two like-tissue samples from the same patient, changes in expression are more reliable than pooling cancer samples and comparing them to pooled normal samples. Although this strategy is very stringent, it is also one of the strengths of our gene expression analysis.

Gene expression data was downloaded from the TCGA database (`https://cancergenome.nih.gov/`). Gene expression data for cancerous tissue and normal tissue in the same patient was available for 114 patients: level-3 HiSeq RSEM gene-normalized RNA-seq data was obtained for 1,104 cancerous and 114 normal samples for 20,530 genes. However, some genes had zero values for one or more patients as a result of either there being no transcripts sequenced, or transcripts missing for a particular gene. Consequently, different methods were attempted to resolve these issues, some of which were not included in the final analysis.

4.2.1 Initial analysis

The initial data analysis workflow for identifying genes down-regulated in breast cancer removed genes with some (84 or more) expression values of zero. The TCGA dataset for breast cancer contains normal and cancerous gene expression data for 20,531 genes in both cancerous and normal tissue across 114 patients. Since the TCGA gene expression datasets for breast cancer are incomplete (i.e. there are numerous genes that do not have 114 expression values for all patients), there were instances of genes with zero values for either cancerous and/or normal samples certain from patients. There were 6,575 genes that had gene expression values of zero for between 1 and 114 patients from cancerous and/or normal

tissue datasets. As a preliminary step, approximately 1,000 genes with zero values for all 114 patients were removed from the analysis entirely due to the lack of data. However, some genes still had zero values for the majority of patients, which would invalidate down-stream statistical analyses. As a result, in order to have sufficient data to complete valid statistical tests (whether parametric or non-parametric), a minimum of 30 non-zero gene expression values were required for each gene. Therefore, 18,654 genes out of 20,531 genes total with at least 30 non-zero values were identified for further analysis in the initial analysis method.

When using statistics to analyze large datasets, it is important to use an appropriate test that reflects the normality of the data. The Shapiro-Wilk test for normality tests datasets to determine whether or not they follow a normal distribution. In order to maintain stringency in our analysis, the Shapiro-Wilk test was conducted on the expression values of the 18,654 genes that had at least 30 non-zero values. The results of this test showed that the gene expression data did not have a normal distribution even though the sample size would have been large enough to make the assumption of normality [28]. Therefore, a non-parametric test was used to analyze this data as it is a more appropriate statistical method for the data.

In spite of the above method of maintaining some zero values in the dataset, by evaluating samples with 30 to 114 paired values per gene (where there is a non-zero value for both cancer and normal tissue), there was the potential for misrepresentation in some genes compared to others, which could lead to potential false positive results. For example, geneA could be found to be down-regulated in breast cancer with a sample size of 30, but in reality if we had genetic data from 114 patients we would see no statistical change in gene expression. As a result, this workflow was not used to generate the breast cancer specific subnetwork.

4.2.2 Final analysis

To improve upon the initial analysis, and to ensure that the results moving forward were statistically sound, any gene for which there was one or more expression value of zero in either cancerous or normal tissue were removed in the final version of the analysis. For example, if there is a zero value in the cancer data set for geneA, then that gene is removed from the analysis regardless of whether it has a zero value in the dataset for normal tissue. This filtering procedure resulted in analyzing data for 13,938 genes out of the 20,531 genes

in the TCGA breast cancer dataset.

Once zero values had been accounted for and dealt with accordingly, the Shapiro-Wilk test was conducted again on the remaining 13,938 genes. The Shapiro-Wilk test showed that although more than 30 samples were present for each gene, the expression values were not normal and a non-parametric statistical test was required. The Wilcoxon signed-rank test was selected as the non-parametric statistical test for the paired data to determine if the median gene expression between the cancer sample and the normal sample was statistically different ($P < 0.05$). Since greater numbers of statistical inferences made from a dataset increases the chances of an error in any of those inferences, it was imperative to account for multiple hypotheses.

4.2.3 Multiple hypothesis testing

Two types of multiple hypothesis corrections were considered for the data analysis, the Bonferroni correction and the Benjamini-Hochberg procedure. The Bonferroni correction corrects for family-wise error rate (FWER), whereas the Benjamini-Hochberg procedure corrects for false discovery rate (FDR)[118].

FWER represents the probability of making a Type I error (false discovery) when testing multiple hypotheses using the same data and reduces the probability of making a type I error with respect to the results returned as positive compared to the original P-value. In contrast, FDR conceptualizes the rate of Type I errors when conducting multiple comparisons. The Bonferroni correction is one of the most strict corrections for multiple hypotheses. Adjusted P-values are calculated by dividing each critical value by the total number of hypotheses (n) being tested (e.g. P-value/n). This procedure affects all critical values equally, which can lead to false negative results because it is so strict. Also the Bonferroni corrections results in a group of hypotheses where α (the allowable percentage of false positive results e.g. 0.05) results are expected to be false positives. Furthermore, the more tests (or the larger n), the greater the effect the Bonferroni correction will have on P-values. In contrast, the Benjamini-Hochberg procedure affects P-values based on their rank (e.g. (P-value's rank/n) * α) where α is the probability of making a Type I error). FDR procedures, such as the Benjamini-Hochberg procedure, are less stringent when it comes to false positives in the data; however,

they have greater statistical power. As a result, for our Wilcoxon signed-rank test on the TCGA data we used the Benjamini-Hochberg procedure, choosing increased statistical power at the cost of potential false positives in the results. In contrast, the Bonferroni correction was used for the Shapiro-Wilk tests for normality on the cancer and non-cancer datasets. It was selected because it is a more stringent statistical test and we wanted to ensure that normality was not being incorrectly assumed.

4.2.4 Determining gene down-regulation in breast cancer

In order to determine if a gene was down-regulated in breast cancer, two analyses had to converge: the Wilcoxon signed-rank test results and the evaluation of gene expression ratio in breast cancer. To determine the fold change of gene expression between the cancerous and normal tissue, the ratio of gene expression was calculated (cancer/non-cancer) and the log-base-2 of the ratio was used as the fold change in gene expression in breast cancer. Once the fold change was calculated and Wilcoxon signed-rank test P-values were corrected for multiple hypotheses, a cut-off value was needed to determine which genes were statistically down-regulated in breast cancer.

We evaluated several potential cut-off values: 2-fold, 3-fold, 4-fold, 5-fold, 8-fold, and 16-fold representing biological significance and one standard deviation representing statistical significance. Though all of these cut-off values are technically valid, it was appropriate to select the cut-off value that most accurately addresses the following question: which genes are down-regulated in breast cancer tissue compared to normal tissue? Since we are asking a biological question, a cut-off that represents an accurate biological significance threshold should be used. As the fold change increases the results become more stringent and the number of significant genes decreases. Biologically a higher fold change means that genes have a greater change in expression when comparing cancerous to normal tissue. However, the aim was to identify genes that are down-regulated, not genes that have large differential expression in breast cancer. As a result, a lower fold-change threshold makes more biological sense. Since a 2-fold cut-off represents instances where a gene in breast cancer has half of the expression observed in healthy breast tissue, it was determined that this cut-off was an accurate representation of the biological question. Consequently, genes that were down-

regulated at a 2-fold or more cut-off in breast cancer and had a P-value (after adjustment) < 0.05 were considered to be down-regulated in breast cancer.

4.2.5 Generating the breast cancer specific subnetwork

To give the SL interactions in HYGIN context in cancer, the results from analyzing the TCGA breast cancer data were used to generate a breast cancer specific subnetwork of HYGIN. The breast cancer subnetwork was generated by extracting the genes that were down-regulated in breast cancer and all of their SL interactions from HYGIN. This subnetwork contains the potential 130 SL interactions (edges) that are specific to breast cancer between 130 genes (nodes) total; 15 genes that are down-regulated in breast cancer and 115 genes that have SL interactions with them.

4.3 Validating HYGIN

HYGIN was validated *in silico* using the Synthetic Lethal Database (SynLethDB; July 12, 2017) [45] and a series of random networks. The SL interactions in the SynLethDB are a combination of computationally predicted SL interactions, and SL interactions that have been identified through large- or small-scale *in vitro* assays. Unfortunately, the text mining algorithm for SynLethDB did not accurately identify SL interactions from the literature. As a result they were all excluded from our validation procedure. Using the remaining information in the database, we identified 11 SL interactions from HYGIN that have been identified by previous studies. Eight of these interactions were computationally identified by DAISY [58], and 3 were identified by SynLethDB through analysis of data in DECIPHER [36].

To determine if the 11 interactions identified by previous studies were statistically significant or if they were the result of random chance, we generated 100,000 iterations of a random network that mimics the original yeast interaction network. By generating multiple random networks we can calculate the number of times that there were 11 or more SL interactions in common with HYGIN that were the result of random chance. The random networks contained the same number of nodes as there were yeast genes (5,587) and the same number of edges as there were yeast SL interactions (132,312) in the yeast SL interaction network. Al-

though there were approximately 550,000 negative genetic interactions in The Cell Map data, only statistically significant entries were used for the generation of HYGIN and subsequent analyses ($P < 0.05$ and $\epsilon < -0.2$, where ϵ is the genetic interaction score used by Costanzo *et al.* [23]). They were converted to a humanized genetic interaction network using the same procedure as for HYGIN, and evaluated for SL interactions in common with the SynLethDB. Random networks and comparison were completed using the Python library NetworkX.

4.4 Verifying one-to-one ortholog mapping of FBXW7

FBXW7 was selected as a target of interest for future analysis not only because it is a known tumor suppressor, but because it is highly connected in the breast cancer subnetwork. With 32 potential SL interactions that can be exploited as therapeutic options, FBXW7 is an ideal candidate for further testing *in silico*, *in vitro*, and *in vivo*. Since one of the major motivations for using one-to-one yeast-human ortholog mapping was the lack thereof in the previous work by Deshpande *et al.*, it was important to verify the one-to-one mapping for the gene of interest. To confirm that there were no additional yeast-human orthologs of FBXW7, the InParanoid database [109] was searched twice: the human genome was searched for the protein MET30 (the yeast protein that is orthologous to FBXW7 in humans) to ensure that there was only one ortholog for the yeast MET30 protein, and for CDC4 (a common synonym for FBXW7 in humans) to ensure that there were no additional orthologs for FBXW7 in the database under its alias CDC4.

4.5 Gene expression in other cancers

A recent study showed that on average 43% of cancers from a specific tissue type were more similar to tumors from a different anatomical site than they were to tumors of the same tissue of origin [24]. Although generally cancer is diagnosed and treated based on the tissue or origin, there is more and more evidence that supports the theory that genetically cancers from different tissues can be more alike than not [53, 61]. As a result, the 15 genes that were in HYGIN and identified to be down-regulated in breast cancer were also analyzed in the

24 types of cancer in the TCGA database. For each of these 24 cancer types, pooled gene expression for cancer tissue was compared to pooled gene expression for normal tissue. Since we were evaluating these cancers for down-regulation, only cancer types where more than 7 of the 15 genes had lower expression in cancer than in normal tissue were reported.

4.6 Drug data analysis

In an attempt to validate the SL interactions of FBXW7 *in silico* we used cancer cell line drug sensitivity data and cancer cell line gene expression data. We performed an *in silico* drug sensitivity assay using data from the cancerRXgene database which contains drug response data for many cell lines [137]. Using cell line gene expression data all cell lines were evaluated and divided into two groups: those with high FBW7 expression, and those with low FBXW7 expression. The rational was that cell lines with low FBXW7 gene expression that are susceptible to a drug provide support that the given drug is a potential therapeutic for targeting the SL partners of FBXW7. In order to show this, the IC50 values for each group (high FBXW7 and low FBXW7) were used to generate a P-value (Mann-Whitney U test). The survival percentage data from the cancerRXgene database for each of 265 drugs at different concentrations was the basis for the dose-response curves. Out of 265 drugs, three were selected for further analysis based on the difference in susceptibility between the FBXW7 high and FBXW7 low cell lines: Cabozantinib (P-value = 0.033), Selumetinib (P-value = 0.006), and NSC-207895 (P-value = 0.099).

Chapter 5

Results

5.1 Construction of HYGIN predicted 10,419 potential SL interactions in humans for 1,009 human genes

Previously, a yeast genetic interaction network, The Cell Map, was described [23]. Using this information and a strict one-to-one ortholog mapping from yeast to human from InParanoid [109] (Supplementary Table B.1), we generated a refined network: HYGIN (Figure 5.1). HYGIN contains all of the yeast orthologs of human genes and their predicted SL interactions. Of the approximate 550,000 negative genetic interactions in yeast (Costanzo 2016; `http://thecellmap.org/costanzo2016/`, updated May 2016) we evaluated only statistically significant interactions with a P-value < 0.05 that were also strong negative interactions where $\epsilon < -0.2$. Using InParanoid mapping from yeast to human, the yeast network was reduced to statistically significant strong negative interactions that exist between the human orthologs of these yeast genes. The resultant HYGIN contains 1,009 genes and 10,419 proposed SL interactions (Supplementary Table B.2). Topology of this network combined with our Gene Ontology (GO) Slim terms (Supplementary Table B.3) shows dense clustering between genes involved in DNA damage and repair pathways and cell cycle regulators; or RNA processing and ribosome biogenesis and translation components, suggesting we have recovered meaningful humanized genetic interactions (Figure 5.1, Figure A.1). Since there are only 1,266 yeast genes that map to exactly one human ortholog, it is not surprising that the total number of human genes in the humanized network is less than that in the starting yeast network, as only a fraction of these genes have human orthologs.

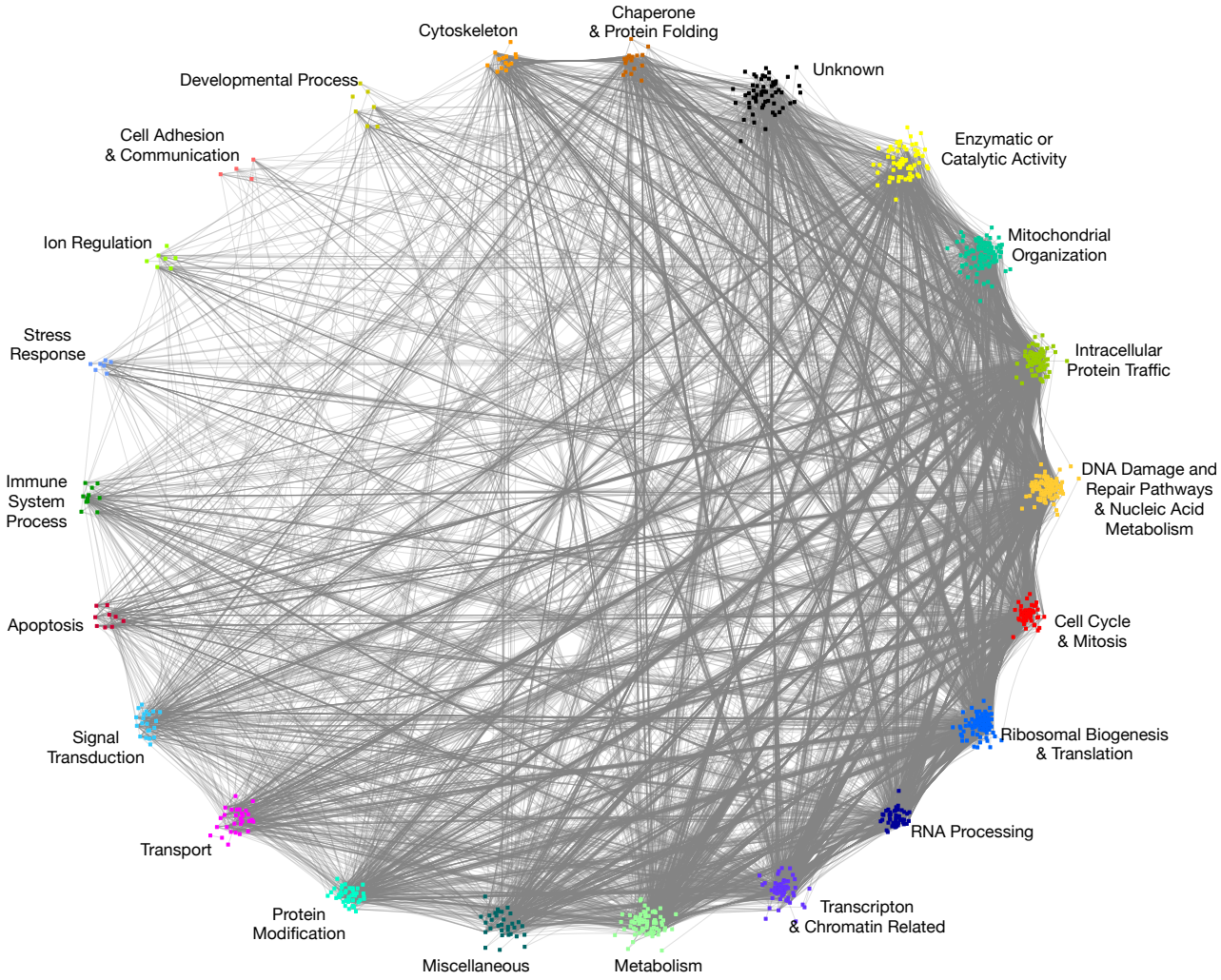


Figure 5.1: HYGIN network. Nodes are clustered around the perimeter of the network and coloured by GO slim terms which represent different biological processes. Each node represents one of the 1,009 human genes in the network. An edge between two genes represents a SL interaction translated from yeast. Denser areas of the network indicate a higher number of SL interactions between GO slim terms, while sparser areas of the network indicate fewer interactions between GO slims. Note that some areas of the network are sparser due to fewer genes representing a specified GO slim term and other areas of the network are over-represented increasing the number of edges in the network.

5.2 Validating the HYGIN network using SynLethDB

In order to validate the interactions in the HYGIN network *in silico*, the SL interactions in HYGIN were compared to the SynLethDB. The SL interactions in the SynLethDB are a combination of computationally predicted SL interactions, and SL interactions that have been identified through large or small-scale *in vitro* assays. Unfortunately, the text mining algorithm for SynLethDB did not accurately identify SL interactions from the literature. As a result they were all excluded from the validation of HYGIN. Of note, 6 SL entries in the SynLethDB identified through text mining were found to have some error or discrepancy (Table 5.1). Some of the errors were minor. For example, an SL interaction between *PSMD14* and *PSMD4* was in the SynLethDB despite the SL interaction being between *RPN11* and *RPN10* in the source manuscript. (RPN11 and RPN10 are synonyms for PSMD14 and PSMD4, respectively.) Other errors were major. A manuscript identifying yeast SL interactions was included in the SynLethDB and said to have human SL interactions; a manuscript that did not contain the words "synthetic", "lethal", or *POLD1* was included in the SynLethDB and recorded as having two SL interactions between *POLD1* and other genes; and most notably, a manuscript whose only reference to SL was in one of the references was identified by the SynLethDB to have 6 SL interactions. The data regarding these errors has been summarized and included as a table (Table 5.1), and the authors of the database were informed about the discrepancies and were eager to address them in the next version of SynLethDB.

Using the information in the database (except for text mining entries), we identified 11 SL interactions from HYGIN that have been identified by previous studies. Eight of these interactions have been computationally identified by DAISY [58], and 3 were identified by SynLethDB through analysis of data in DECIPHER [36]. The genes involved in these 11 interactions are summarized in Table 5.2. Since the number of SL interactions in common was low we wanted to test to see if these interactions could be the result of random chance or not. To do this we performed 100,000 iterations of the yeast-human ortholog mapping process starting with a random network that mimics the original yeast interaction network. It was found that there were 0 iterations that resulted in a humanized network having 11 or more SL interactions in common with the SynLethDB. However, there were random networks

PubMed ID	Description of error	SL interactions in SynLethDB
23651857	LSM2 does not appear in the manuscript and there is no direct mention of synthetic lethal interactions in the manuscript itself. The only reference to synthetic lethal interactions is one of the references: Luo J, Emanuele MJ, Li D, Creighton CJ, A1:G7 MR, Westbrook TF, Wong KK, Elledge SJ 2009. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 137: 835–848	*GSC LSM2* *LSM2 MYC* *LSM2 PHF5A* *GSC MYC* *GSC PHF5A* *MYC PHF5A*
16135799	FEN1 and FBXO18 do not appear in the manuscript. The manuscript is about yeast synthetic lethal interactions, not human synthetic lethal interaction.	*FBXO18 RAD51* *FBXO18 FEN1* *FEN1 RAD51*
25139395	POLD1 is not mentioned in the manuscript. Neither are the terms "Synthetic" nor "Lethal".	*POLD1 RAD51* *RAD51 RELA* *POLD1 RELA*
23980094	The only two valid synthetic lethal interactions in this manuscript are ASPSCR1 and PSMC2, and SMARCB1 and PSMA4. The other interactions for this manuscript listed in the database are just every possible combination of these two genes.	*ASPSCR1 PSMA4* *ASPSCR1 PSMC2* *PSMA4 PSMC2* *PSMA4 SMARCB1* *PSMC2 SMARCB1* *ASPSCR1 SMARCB1*
20941496	Valid SL interaction, although synonyms of the genes in the source manuscript are used: PSMD14 is a synonym for RPN11, and PSMD4 is a synonym for RPN10 where neither PSMD14 nor PSMD4 appear in the manuscript.	*PSMD14 PSMD4*
25334017	This manuscript is about embryonic lethality, not synthetic lethality. CDC45 only appears in the manuscript 3 times (once in the abstract and twice in the introduction), and is not described in the manuscript as involved in a synthetic lethal interaction.	*CDC45 GINS1* *CDC45 GINS3* *GINS1 GINS3* *GINS3 TAP2* *CDC45 GINS4* *GINS3 GINS4* *GINS4 TAP2* *CDC45 TAP2* *GINS1 TAP2* *GINS1 GINS4*

Table 5.1: List of SL interactions identified in the SynLethDB that were found to have errors. "PubMed ID" is the PubMed accession number of the source manuscript.

GeneA	GeneB	PubMedID	Source
KRI1	NOP56	25171417	DAISY
NCAPG	NCAPH	25171417	DAISY
TAF5	YEATS4	25171417	DAISY
DLD	SUCLA2	25171417	DAISY
CHAF1B	FEN1	25171417	DAISY
KIFC1	NCAPD2	25171417	DAISY
MCM6	WDHD1	25171417	DAISY
LIG1	RAD51	25171417	DAISY
GTF2H1	POLE	2667985	DECIPHER
PMS2	POLD1	2667985	DECIPHER
POLD1	POLE	2667985	DECIPHER

Table 5.2: List of SL interactions identified in HYGIN and the SynLethDB.

Number of Edges	Number of Networks
1	364,951
2	209,841
3	79,255
4	22,415
5	5,094
6	943
7	140
8	22
9	7
10	1

Table 5.3: A list of the number of edges in common between the SynLethDB and the 100,000 random networks used for validation. The number of edges column represents the number of edges in common between the two networks, and the number of networks column represents the fraction of random networks (out of 100,000) that contain that number of edges in common with the SynLethDB.

that had between 1 and 10 edges in common with the SynLethDB (Table 5.3). As a result, we can say with high confidence (P-value < 0.00001) that identifying these 11 interactions in common between HYGIN and the SynLethDB were not the result of random chance. Using this approach, we were able to show that the interactions in the network are statistically significant when compared to random networks.

5.3 Identification of down-regulated genes in breast cancer and building a breast cancer-specific SL interaction network

To make HYGIN more applicable for human cancer therapeutics, we focused on breast cancer and developed a breast cancer specific subnetwork. Exploiting any yeast SL interaction may not be beneficial if neither of the genes is altered in cancers. As a result, in order to identify those interactions in HYGIN that are relevant to breast cancer, we used TCGA data (`https://tcga-data.nci.nih.gov/tcga/`) to identify genes that are down-regulated in breast cancer. Although there are gene expression data for over 1,000 breast cancer patients available in TCGA, because gene expression is a relative measurement, we chose to use only those tumor samples that had matching normal gene expression data. That is, we used data only from patients where gene expression results were available for both cancer and normal tissue. This stringent approach was followed to eliminate those genes that are down-regulated in cancer, but whose expression in normal tissue is also low. In order to determine if genes were statistically down-regulated in breast cancer, we first had to determine which statistical test to use. The Shapiro-Wilk test for normality was used on the initial data and results showed that values from 80% of genes in the cancer dataset had normal distributions, compared to 56% of the genes in the dataset for normal tissue. These results indicated that despite being able to assume that the data followed a normal distribution [28], a non-parametric test was more appropriate for analyzing the data.

Our stringent procedure removed genes with any zero values for calculating gene up- and down-regulation in breast cancer. This was a very stringent method of analyzing the data,

and resulted in the removal of 6,593 genes from the initial dataset of 20,531 genes. The resulting 13,938 genes had non-zero values for both the cancerous and normal datasets for all 114 patients. Although this stringent cut-off potentially removed genes that are statistically down-regulated in breast cancer, it ensured a consistent sample size for all genes and reduced the potential for false positive results. Given the change in dataset, the Shapiro-Wilk test for normality was run again: 75% of expression values for cancer genes were normally distributed, compared to 46% of normal genes. Therefore, to be consistent across all data, the Wilcoxon signed-rank test was used. It determined statistical differences in the median gene expression values for each gene between the cancer and normal tissue.

For each patient, gene expression data for cancerous and normal tissue was compared to generate a log-2 ratio of cancerous to normal gene expression. By applying a 2-fold cut-off and using genes with statistically different medians (P-value after adjustment < 0.05), 1,745 genes were identified to be down-regulated in breast cancer (Figure 5.2, Supplementary Table B.4). Of these 1,745 genes that are statistically down-regulated in breast cancer, 181 were previously identified in the tumor suppressor gene database (TSG; `https://bioinfo.uth.edu/TSGene/`) as potential tumor suppressors.

The 1,745 genes that were statistically down-regulated in breast cancer were compared to the HYGIN network, and it was found that 15 genes were in both datasets. These 15 genes included some of the well-established tumor suppressors like *CAT* [35] and *FBXW7* [18]. The breast cancer specific subnetwork was generated using these 15 down-regulated genes and the 115 genes that they have SL interactions with for a total of 130 genes (Figure 5.3, Supplementary Table B.5). The 115 genes are themselves not down-regulated in breast cancer. By targeting the SL partners of the 15 genes that are already down-regulated in breast cancer, novel targeted therapeutics can be developed to provide treatment options for patients. A flow diagram of the methods used to analyze the breast cancer data and generate the breast cancer specific subnetwork is depicted in Figure 4.1. An interesting note is that these 15 genes were also found to be statistically down-regulated in breast cancer when all cancer samples were pooled and compared to all non-cancerous samples (Figure 5.4, Figure A.5, Figure A.6, Figure A.7).

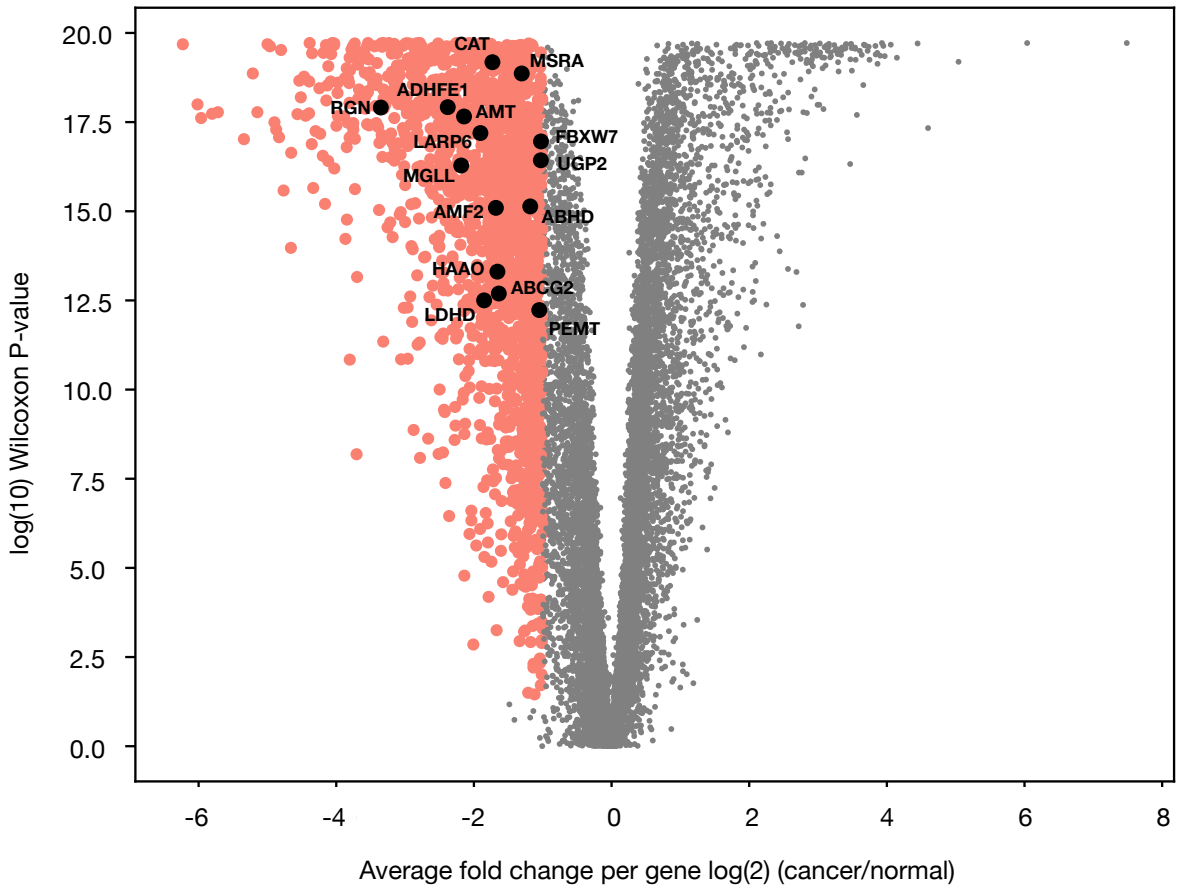


Figure 5.2: Volcano plot of the average change of expression per gene expressed as log(2)[cancer/normal] plotted against the log(10) of the adjusted Wilcoxon P-value. Dots in red represent genes that have an average 2-fold or more decrease in expression. Back dots represent the 15 genes that are also in the HYGIN network.

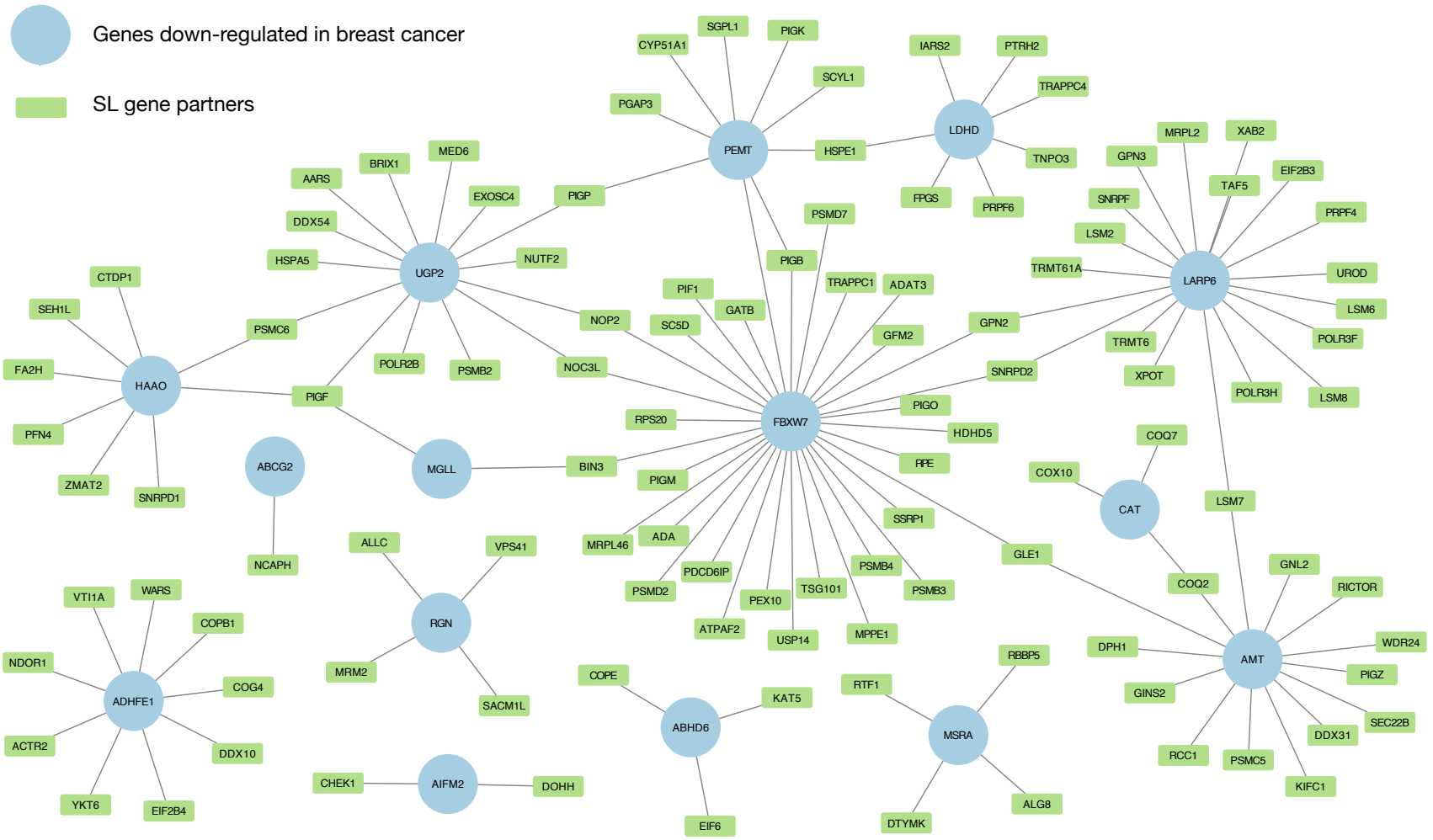


Figure 5.3: Visualization of the breast cancer specific subnetwork generated using 15 genes in HYGIN that were found to be down-regulated in breast cancer (blue circles) and their corresponding SL gene pairs. Edges represent proposed SL interactions between genes, and green nodes represent SL partners of the 15 genes.

5.4 Synthetic lethal interactions of *FBXW7* and their validation

Having identified *FBXW7* as a candidate gene worth exploring further, its yeast ortholog and expression in other cancers was investigated. The yeast gene *MET30* is an F-box protein containing five copies of the WD40 motif and is known to control cell cycle function as part of a E3 ubiquitin ligase complex [119]. According to InParanoid, FBXW7 is the human ortholog of MET30 that shares similar roles in humans [109] (Figure A.2) In fact, *FBXW7* is a well-known tumor suppressor that functions as a substrate-recognition protein within a SCF (SKP/CUL1/F-Box) E3 complex ubiquitin ligase complex, which targets numerous proteins for ubiquitin-mediated proteasomal degradation [130]. *FBXW7* is down-regulated not only in breast cancer, but also in 12 other cancers including colon, liver, lung, and prostate cancers (Figure 5.4, Figure A.3). However, *FBXW7* is also up-regulated in 6 other cancers including three types of kidney cancer, lung adenocarcinoma, and thyroid carcinoma (Figure 5.4, Figure A.3). Although SL interactions are potential targets for cancer therapeutics in some cancers, it would not be appropriate to use this strategy on all tumor types. Therefore, targeting SL interactions of *FBXW7* may have a wide opportunity for clinical application, but must be used in conjunction with genetic testing and patient gene expression data.

In order to validate the novelty associated with the identified SL interactions in breast cancer, we focused on SL interactions with *FBXW7*. Starting with *FBXW7* our analysis identified 32 SL interactions, many of which are proteasome components such as PSMB3, PSMB4, PSMD2, PSMD7, and USP14. Given that FBXW7 is a E3 ubiquitin ligase, these interactions highlight the genetic property of SL relationships where functional coherence is often observed. As cyclin E is a substrate of proteasome degradation, it is interesting to note that the recently published SL interaction between *FBXW7* and *CCNE1* may reflect the interaction between FBXW7 and proteasome components [4]. Our ortholog mapping is one-to-one and unfortunately, cyclin E falls into the"more than one ortholog" category in yeast and "more than one corresponding ortholog" in humans, and as a result was not included in HYGIN.

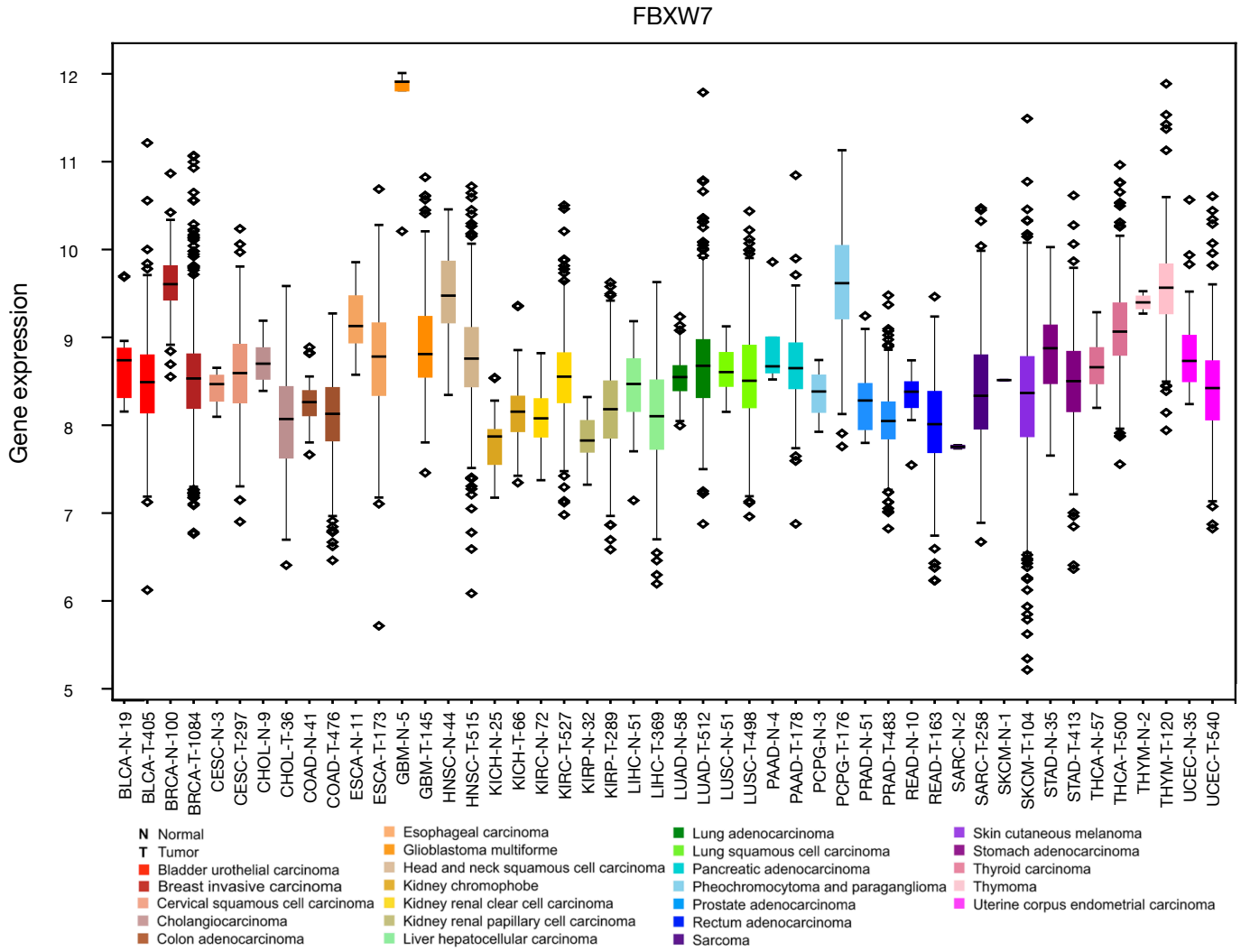


Figure 5.4: Gene expression of *FBXW7* in 24 cancer types and corresponding normal tissue as calculated from RNA-seq data by Expectation-Maximization (log2). The averages of normal and tumor expression of *FBXW7* for the same tissue type are plotted beside each other and colour-coded based on cancer type. Number of patient samples is indicated in the X-axis labels.

Additionally, *FBXW7* and *PEMT* are both separately down-regulated in breast cancer, yet they share a proposed SL interaction in HYGIN. The gene expression ratios for all 114 patients for both *FBXW7* and *PEMT* were evaluated to further investigate this finding. Results showed that 20% of patients have gene expression ratios greater than 2-fold for both *FBXW7* and *PEMT*, and one patient has gene expression less than 4-fold for both genes. This information suggested that the SL interaction between *FBXW7* and *PEMT* is a false positive result since the inhibition of both genes in a SL interaction should lead to cell death.

Our computational prediction suggested a SL interaction between *FBXW7* and *USP14*, a gene that codes for a deubiquitinating enzyme. FBXW7 has been proposed in the degradation of a number of substrates including Aurora B [20, 117]. Although Aurora B kinase is primarily degraded through the anaphase-promoting cyclosome complex (APC/c) [81], negative regulation of Aurora B by FBXW7 plays an important role in Aurora B ubiquitination and degradation [117]. USP14 is a major regulator of the proteasome and one of three proteasome-associated deubiquitinating enzymes, which also affects protein turnover in a substrate-specific manner [66]. Interestingly, a recent study has reported that over-expression of USP14 stabilized and prevented Aurora B degradation through deubiquitination [108]. Furthermore, a FBXW7-Aurora B-p53 negative feedback loop has been suggested [128]. This feedback loop suggests that a loss of *FBXW7* leads to an increase in Aurora B, which phosphorylates p53 and leads to MDM2 enhanced degradation of p53 and ultimately cancer cell growth [128]. While the loss of FBXW7 may already stabilize Aurora B [20, 117], over-expression of USP14 may be necessary for continued stability of Aurora B to maintain cell-cycle progression and cell survival. Thus, consistent with our observation, loss of USP14 when *FBXW7* is down-regulated may destabilize Aurora B, leading to a SL phenotype.

We also predict *FBXW7* to exhibit SL interaction with a member of the pentose phosphate pathway (PPP), ribulose-5-phosphate-3-epimerase (RPE), which catalyzes the reversible epimerization of D-ribulose 5-phosphate to D-xylulose 5-phosphate. Interestingly, previous work has shown that inhibition of the PPP results in decreased proliferation of tumor cell lines [12, 13, 19, 21, 98]. In addition, it has also been shown that the PPP is essential for metabolic network modulation to support tumor angiogenesis as inhibition of VEGFR-2 causes a decrease in PPP flux [126]. Thus, if inhibition of VEGFR-2 may decrease PPP flux, we hy-

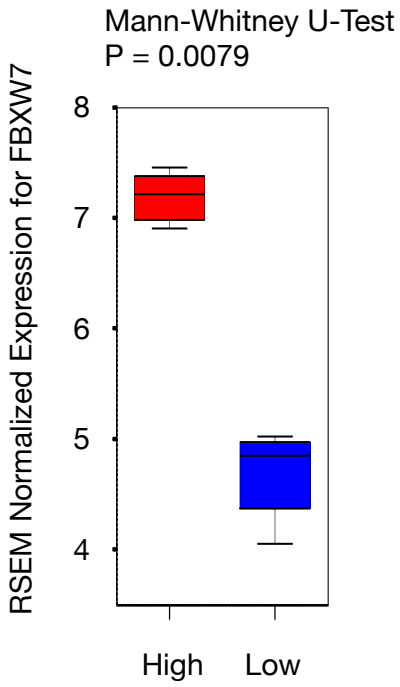


Figure 5.5: Gene expression data for *FBXW7* was extracted from all cell lines studied and cell lines were split into two categories: high *FBXW7* expression and low *FBXW7* expression. Here, a box plot shows *FBXW7* expression in all cell lines; up-regulated (red) and down-regulated (blue). The P-value (0.0079) represents the statistical difference between the two populations: cell lines with either high or low FBXW7 expression.

pothesized that inhibition of VEGFR should mimic the SL interaction between *FBXW7* and *RPE*. To test this idea, we used drug response data from the cancerRXgene database (`https://www.cancerrxgene.org`) and asked if cell lines deficient in *FBXW7* are sensitive to the VEGFR-2 inhibitor Cabozantinib. The cancerRXgene database contains data for 265 drugs and multiple cell lines and was examined to identify compounds that are more effective when used selectively with cell lines that have a low *FBXW7* expression. We found that cell lines with low expression of *FBXW7* were more sensitive to Cabozentanib (P-value = 0.033) (Figure 5.5, 5.6), which supports the idea that inhibition of VEGFR should mimic the SL interaction between *FBXW7* and *RPE*.

Similarly, we also found a SL interaction between *FBXW7* with *ADA*, an adenosine deaminase that regulates cellular levels of adenosine and deoxyadenosine. ADA, along with a few other enzymes (for example HGPRT), are responsible for purine metabolism and are

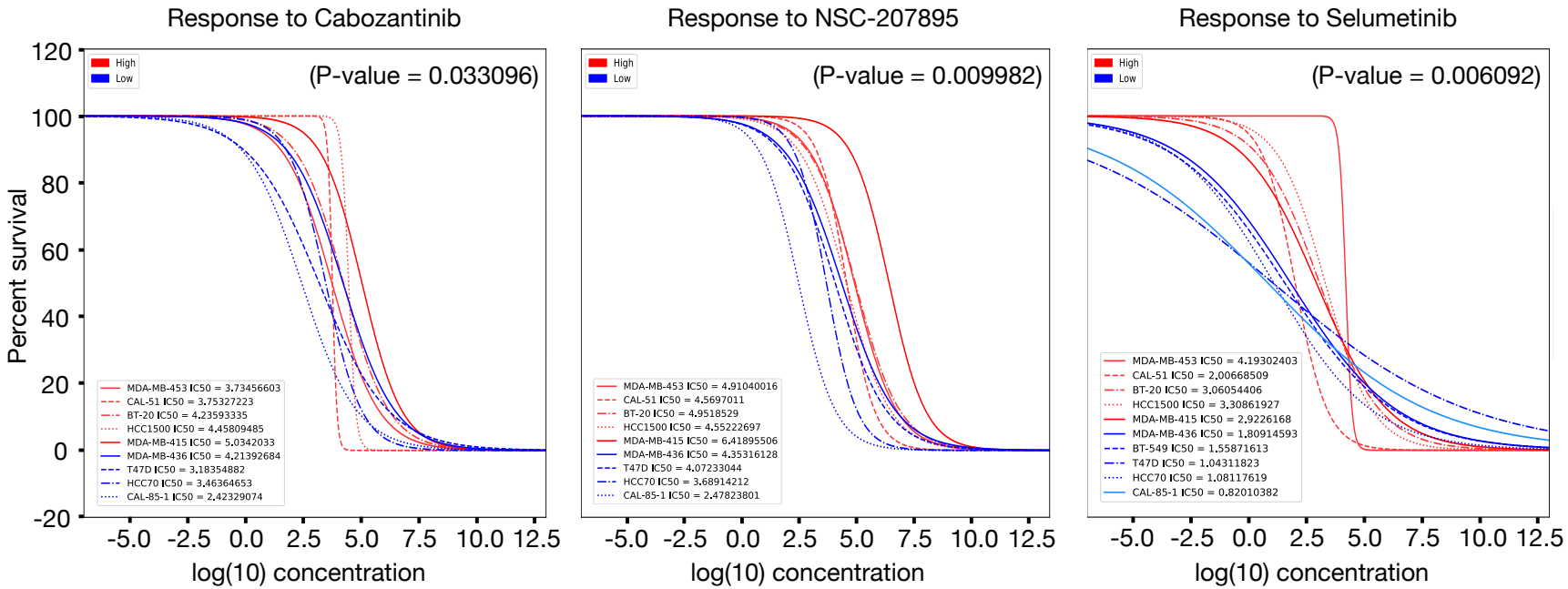


Figure 5.6: Cell line gene expression data and cell line drug data (cancerRXgene database) were evaluated to identify to which drugs low *FBXW7* cell lines are susceptible. Drug kill curves in blue represent cell lines where *FBXW7* is down-regulated, and drug kill curves in red represent cell lines that have up-regulated *FBXW7*. Drug curves for Cabozantinib (left), NSC-207895 (centre), and Selumetinib (right) were identified to have the greatest statistical significance between cell lines with high and low FBXW7 gene expression.

well known targets for cancer chemotherapy [38]. ADA can both degrade adenosine and bind extracellularly to adenosine receptors to function as an allosteric modulator to regulate adenosine. In fact, studies in a human astrocytoma cells have shown that manipulation of cellular purine metabolite concentrations can make these cells more sensitive to apoptosis [39]. While it remains to be explored if the astrocytoma cell line is deficient in *FBXW7*, these studies indicate a potential role for *ADA* as a therapeutic target. Interestingly, previous studies have shown that the expression of *ADA* is induced by growth factors like IGF in a Ras-MAPK pathway dependent manner [134]. To test this idea, we used drug response data from cancerRXgene database and asked if cell lines deficient in *FBXW7* are sensitive to the *MEK* inhibitor Selumitinib (Figure 5.6). Consistent with our SL observation, we found cells with low expression of *FBXW7* to be more sensitive to Selumitinib (P-value = 0.006). Thus, our observation strongly indicates that *FBXW7* can be used as a biomarker to indicate whether to treat patients with drugs like Selumitinib that indirectly affects the expression of *ADA*. Importantly, the availability of ADA inhibitors like 2'deoxycoformycin (Pentostatin) that are already in the clinics as standard chemotherapeutic agents for lymphoid malignancies[83] may also represent a potential option to test our prediction. Indeed, analyses of the drug data also indicates a potential opportunity for the MDM2 inhibitor NSC-207895 in treating FBXW7 deficient cancers (P-value = 0.099, Figure 5.6). Using an MDM2 inhibitor to treat *FBXW7* deficient cancers is interesting because MDM2 is an negative regulator of p53 [64]. Previous research has also showed that *FBXW7* is a transcription target of p53, and that *FBXW7* is dramatically up-regulated when p53 is introduced into p53-defficient cells by adenovirus [62]. Therefore, if we use an MDM2 inhibitor, we would expect an increase in p53 and subsequently an increase in *FBXW7*, which would mean that targeting *FBXW7* SL interactions would be not be effective in this instance. However, there is also evidence that supports FBXW7 mediating the role of P53 in response to DNA damage [72]. As a result, although the *in silico* drug analysis suggests that an MDM2 inhibitor would be a valuable drug for treating *FBXW7* deficient breast cancers, additional *in vitro* analyses are imperative in order to understand what is going on inside the cell.

Chapter 6

Discussion

In this thesis we describe a novel approach to identifying potential SL interactions in breast cancer that uses a yeast SL interaction network, yeast-human gene ortholog information, and patient gene expression data. Yeast is a common model organism and generating a human orthologous network from a yeast gene network provides a major framework to extend yeast genetic interaction data to humans. When generating HYGIN, strict one-to-one yeast-human ortholog mapping was used to avoid ambiguity as there are several instances where a single yeast protein maps to multiple human orthologs, and vice versa. HYGIN is cancer independent as no assumptions about cancer are made while generating the network. As a result, any cancer gene expression data can be subsequently applied to generate cancer type-specific subnetworks of genetic interactions. In this thesis we generate a breast cancer specific subnetwork of HYGIN using breast cancer patient data from the TCGA.

6.1 Network generation

The development of HYGIN was based on previous work by Deshpande *et al.*, but took into consideration some of the problems with their network to generat a more robust and strict network of human SL interactions [27]. Previously, Deshpande *et al.* generated two networks: a complete SL network using all available orthologs in InParanoid, and a cancer-specific network. The first network did not discriminate between any of the yeast-human orthologs in InParanoid and simply applied any available orthologs (whether they were one-to-one, one-to-many, many-to-one or many-to-many) to the yeast data. The second network, used strict one-to-one yeast-human ortholog mapping network and contained cancer-relevant mutation data. This thesis identified several discrepancies in their initial larger network

related to problems with Ensembl identifiers, accession numbers, and gene names of the SL interactions. Initially it was thought that these errors were the result of inconsistent bookkeeping. However, perspective and more information suggested that these mistakes are due to changes in the databases (over the past 8 years) as well as updates to the human genome, including characterization of previously uncharacterized genes and proteins. This brought to light an interesting problem for HYGIN: how to make this research stand the test of time to the best of our ability now? Researchers do not anticipate that entries in a database or a reference table change over time or become deprecated; however, it was something to consider when developing HYGIN: how to ensure that the network would be relevant in the future? As a result, HYGIN has two reliable identifiers for each entry. In addition to the gene name, a UniProtKB identifier is also present for each protein made from that gene. Since UniProtKB is a fundamental bioinformatics database heavily utilized in current research, any deprecated identifiers will be maintained in the database, with pointers to updated identifiers. At the very least, deprecated identifiers will be recorded in the entry for a particular protein, and that protein can still be located using the deprecated identifier as a search term.

The Deshpande *et al.* network also has 48 instances of non-canonical names used in the common name list that contain either '-' or '.' and a version number as part of the name. For example, the network contained the name "06-Mar". It is known that this name can be the result of automatic formatting when "MAR6" is typed into Excel [140]. It is still possible that there are outliers in gene and protein accession lists that were used in this thesis, and that there are minor mistakes in HYGIN. However, steps were taken to ensure that these anomalies were minimized.

It is also worth discussing the differences in ortholog mapping between HYGIN and the Deshpande *et al.* research. For example, in the instance where there are multiple human orthologs for a single yeast protein, a corresponding node in the Deshpande *et al.* network now represents multiple human proteins. If we are attempting to compare their network to previously identified SL interactions, we are left with a challenge. All of the nodes can be "unpacked" such that there is now a node representing each protein and a new edge from each of these proteins to its SL partners. However, this "unpacking" of complex nodes can drastically increase the number of edges and nodes in the network. In the case of the Desh-

pande *et al.* network, the number of edges would quickly double from the addition of these new edges resulting in a new network that has more than 50,000 edges. This would generate a large number of potential falsely identified SL interactions in humans as well. However, the second network generated by Deshpande *et al.* was a cancer-specific one that only used one-to-one ortholog mapping for higher stringency. The potential problems identified in the Deshpande *et al.* work was one of the motivations to use strict one-to-one ortholog mapping for HYGIN.

Furthermore, HYGIN used the same cut-offs ($P < 0.05$ and $\epsilon < -0.2$) as the Deshpande *et al.* network. However, they also used a relaxed cut-off ($P < 0.05$ and $\epsilon < -0.08$) for SL interactions that were identified in reciprocal screens (where query A was crossed with array B and vice veresa [22]). In our work to maintain stringency we maintained the same cut-off ($P < 0.05$ and $\epsilon < -0.2$), irrespective of whether an interaction is identified in a reciprocal screen or not. Their work was compared to HYGIN to determine if there were more interactions identified in our work. Compared to Deshpande *et al.* [27] who identified 1,522 potential SL interactions in humans, we identified 10,419 interactions. Interestingly, there are no SL interactions in common between these two networks. A potential reason that HYGIN contains more SL interactions is that our study uses a more recent version of The Cell Map database that includes interactions between essential and nonessential yeast genes. Another reason is that their aim was to generate a SL gene network involving cancer-relevant mutations in human cells. In contrast, our aim was to generate a SL gene network in a cancer independent context, then apply cancer gene expression data to identify SL interactions in breast cancer. As a result, the SL interactions in their network have cancer relevant mutations, whereas no assumptions about cancer were made in the generation of HYGIN.

6.2 Statistical analysis

There was trial and error involved in analyzing the TCGA data for the breast cancer sub-network. Similar to the initial HYGIN, the initial analysis of the breast cancer data lacked statistical validity. As a result, in order to have more confidence in our results, we applied

strong statistical methods to the data which yielded a more concise breast cancer subnetwork. The final methods for data analysis involved removing genes with any zero expression values, evaluating an expression ratio for each gene for each patient, statistically determining if the expression was down-regulated (Wilcoxon signed-rank test), multiple hypotheses adjustment (Benjamini-Hochberg procedure), and setting a cut-off value for genes identified to be down-regulated in breast cancer (biological cut-off of 2-fold loss of expression).

An interesting result was observed while evaluating the breast cancer gene expression data to determine genes that were down-regulated in breast cancer. The expression of the 15 genes that were identified to be in HYGIN and down-regulated in breast cancer were also analyzed in 24 different cancer types. This analysis used pooled normal versus pooled cancerous tissue in order to determine if a gene was statistically up- or down-regulated in the other cancer types. Though previously it was discussed that a method using pooled data is less stringent (Section 6.1), it is worth noting that in breast cancer, even when all 1,084 cancer samples were pooled and compared to all 100 normal samples, all 15 genes were still statistically down-regulated. This could largely be due to sample size, and caution should be taken when using this method for other cancer types that have fewer normal tissue gene expression datasets. Using this data as a starting point to explore these 15 genes and their SL interactions as methods of treatment in additional cancer types is promising.

6.3 Network validation

When validating HYGIN we used previously identified SL interactions in humans and took an *in silico* approach to validation rather than an *in vitro* approach. Though *in vitro* work is still required to successfully validate that the interactions in HYGIN are biologically relevant, the *in silico* validation provides support for our results. Of the nearly 35,000 SL interactions recorded in the SynLethDB, 5,830 were identified with DAISY, an *in silico* approach to the identification of SL and synthetic dosage lethal (SDL) interactions in humans [58]. DAISY uses both patient and cell line data to computationally predict SL and SDL interactions using three inference strategies. Yet only 8 of these SL interactions identified by DAISY were found in HYGIN. One explanation for this is that HYGIN is cancer independent. As a result,

relying solely on yeast orthologs in humans and previously identified SL interactions in yeast, human genes without yeast orthologs were excluded from our network. In contrast, DAISY evaluated existing cancer gene expression data, cancer mutation data, and cancer cell line data to computationally predict SL interactions. When comparing the number of genes in each of the studies, HYGIN analyzes 1,009 genes and DAISY more than 20,000 genes. This difference in sample size alone could account for the lack of overlap between the two studies. However, it is interesting that DAISY did not predict more interactions between human genes with known yeast orthologs. One reason for this could be the focus on cancer gene expression, mutation, and cell line data that drive DAISY towards potential cancer-relevant SL interactions and away from SL interactions involving non-cancerous genes.

There were also 1,020 SL interactions as a result of analyzing the shRNA data in the DECIPHER database [45]. Since the DECIPHER database does not contain the genetic information of healthy individuals, SL interactions predicted from the analysis of bi-specific shRNA screens also have a tendency to identify SL interactions involving known cancer oncogenes and tumor suppressors. Therefore, it is not surprising that there were only 3 SL interactions identified in both HYGIN and DECIPHER given the cancer independent nature of HYGIN, and the small number of patient-specific SL interactions identified in DECIPHER.

The 11 interactions from two separate sources provide evidence that there are SL interactions in HYGIN that have been previously identified in humans. Although it would be arduous to exhaustively test all 10,419 SL interactions in the network *in vitro*, these 11 suggest that HYGIN has identified a potential 10,408 SL interactions between 1,009 genes that have not previously been reported in the SynLethDB.

6.4 Breast cancer subnetwork

The breast cancer specific subnetwork contains several SL interactions involving genes that were determined to be down-regulated in breast cancer patients. The subnetwork, and ultimately this subset of SL interactions, are an ideal starting point for *in vitro* and *in vivo* validation studies and suggest novel targeted therapeutic strategies. Although there are only 130 interactions in the breast cancer subnetwork, validating all these interactions would be

a time- and resource-consuming task. In contrast, additional *in silico* analyses is a more cost- and resource-effective approach to identify SL candidates to test *in vitro*. By extracting 15 genes that were found to be down-regulated in humans and their SL partners, we can target their SL partners with known or novel drugs to identify new treatment strategies on a patient-by-patient basis.

The breast cancer specific interaction network produced 130 SL interactions between 15 genes down-regulated in breast cancer and their 115 SL partners. We further analyzed the 32 genes that have proposed interactions with the tumor suppressor *FBXW7* and found that five of those genes have previously identified drug inhibitors. The remaining genes that interact with the 15 down-regulated genes in the breast cancer network are also potential targets for further studies.

FBXW7 and *PEMT* are both down-regulated in breast cancer and they share a proposed SL interaction in HYGIN. The gene expression ratios for all 114 patients for both *FBXW7* and *PEMT* were evaluated to further investigate this finding. Results showed that 20% of patients have gene expression ratios less than 2-fold for both *FBXW7* and *PEMT*, and one patient had gene expression less than 4-fold for both genes. There are two possibilities that could explain this result. First, it is possible that at some point in evolutionary time this SL interaction was valid. However, the cancer cells overcame the SL dependency of these two genes in humans, and what we expect to manifest as cell death results in cell survival. Second, it is possible that this result is a false positive. In light of a study that showed that there was only 23% overlap in SL interactions between two species of the same genera (*Saccharomyces cereviseae* and *Saccharomyces pombe*) [29], SL relationships are lost over evolutionary time. Thus, we are expecting to see a loss of some SL interactions in our network as a result of a larger species gap between *Homo sapiens* and *Saccharomyces cerevisieae*. Ultimately, this result highlights the need for biological testing of predicted SL interactions prior to beginning therapeutic trials. Though the HYGIN network is based on previous valid research methods, human tissue behaves differently than predicted results.

The breast cancer specific network contains interactions that are valuable starting points for SL research moving forward and provides numerous SL interactions for the scientific community to study. The 15 genes that were identified to be down-regulated in breast cancer

were also assessed in 24 other cancer types. Of the 15 genes, 11 of them were statistically down-regulated in more than 15 cancer types including breast cancer (Figure A.4, A.5, A.6, and A.7). Of note, *HAAO*, *MGLL*, and *UPGE* were down-regulated in 18 cancer types; *ADHFE1*, *CAT*, and *MSRA* were down-regulated in 19; and *RGN* was down-regulated in 21 different types of cancer. This information suggests that SL interactions identified in breast cancer for these genes would be extremely valuable to explore in other cancers as well. We predict that *in vitro* and *in vivo* studies for these genes and their SL partners in cancer models could lead to novel treatments for breast cancer that could be extended to other cancer types.

6.5 Limitations of this study

Although in this thesis the HYGIN network was generated and validated, it is important to address some of the limitations of this approach and how they impact this work moving forward. First and foremost, it is important to acknowledge that the initial yeast dataset represents a small portion of the potential interactions in the human genome. This was made evident through the ortholog mapping, since only 1,009 yeast genes have strict one-to-one human orthologs. As a result, we are essentially comparing approximately 16 - 20% of the yeast genome to 5% of the human genome. This reduction in sample space is extremely advantageous for saving time in the laboratory, but can result in overlooking numerous potential SL interactions. Additionally, using strict one-to-one ortholog mapping also affects the number of genes present in HYGIN. There are more than 1,000 additional yeast genes that have one-to-many or many-to-one orthologous relationships with human genes. The TCGA analysis eliminated any genes that had a non-zero value for any patient (either cancer or normal datasets) which resulted in approximately 30% of human genes being excluded in the breast cancer down-regulation analysis. Furthermore, using stringent P-value cut-offs and a 2-fold cut-off for down-regulation also potentially excluded a number of potential genes and their SL interactions from the analysis. Therefore, although this analysis identified potential candidates for patient specific therapeutics, it very likely missed many other potential relevant SL interactions, and there is room for expansion and improvement moving forward.

Chapter 7

Future Work and Conclusions

7.1 Future Work

Throughout the work for this thesis a large amount of data was analyzed and used to generate HYGIN and the breast cancer specific subnetwork. However, only a small portion of the results were analyzed in great detail: *FBXW7* and its SL partners. As a result, the remaining 14 genes down-regulated in breast cancer and their 105 SL partners, the remaining 1,730 genes down-regulated in breast cancer, and the remaining 10,289 SL interactions in HYGIN were not analyzed in detail. This section discusses the potential future analyses of the data generated through this project, potential caveats of future analyses and benefits to patient care as a result.

7.1.1 Analyzing the 15 genes down-regulated in the breast cancer subnetwork

In addition to evaluating the role of *FBXW7* and its SL partners in breast cancer, future work could also apply the same analyses to the remaining 14 genes that are down-regulated in the breast cancer subnetwork. The range of SL partners for these 14 genes is 1 to 19, with 7 genes having fewer than 5 SL partners (Figure A.3). Analyzing the 7 genes that have more than 5 potential SL partners that can be used as drug targets (*LDHD*, *HAAO*, *ADHFE1*, *PEMT*, *AMT*, *UGP2*, *LARP6*) could lead to additional treatment strategies for breast cancer patients. Moreover, patients who have more than one of these genes down-regulated in their tumor could benefit from combined drug therapy, which could increase the effectiveness of treatment and recovery. The same method that was used to identify *FBXW7*

deficient cell lines that were susceptible to known chemotherapeutics can also be applied to these 14 genes. This analysis method would provide a similar amount of data for each of these genes, which could lead to repurposing current therapeutics on a patient-by-patient basis. This data could also provide insight into the mechanism behind some of the SL interactions that were identified. Ultimately, however, this information could provide insight into which genes are the best candidates for *in vitro* and *in vivo* studies moving forward by highlighting genes that have the most clinical relevance and potential for success. Given that there are 130 SL interactions in the subnetwork, narrowing in on one or two genes to study *in vitro* or *in vivo* would be resource- and time-efficient, especially if there was sufficient evidence to support drug susceptibility in cell lines deficient in these genes.

7.1.2 Expand methods to other cancer types

Since the TCGA database has the same format for all of the cancer types and the Vizea-coumar research group routinely downloads the most up-to-date version of all of the TCGA data, running the same analysis on the remaining 23 cancer types in the TCGA is very straightforward. However, the biggest challenge arises from the lack of normal data for some of the cancer types. For this study, we were fortunate enough to be able to use the dataset from the TCGA database with the most paired samples (114). Other cancer types in the TCGA database have between 1 and 72 paired cancerous and normal samples. In instances where there are 72 normal samples the argument could be made that sample size is large enough for statistical validity. However, in instances where there are only one or two normal samples, it is difficult to analyze the data with statistical confidence. As a result, analyzing the remaining cancer datasets can be prioritized based on the number of paired samples available. The breast cancer dataset had the largest number of paired samples, and the next top 5 datasets are: kidney renal clear cell carcinoma (KIRK-US): 72 samples, thyroid carcinoma (THCA-US): 58 samples, lung adenocarcinoma (LUAD-US): 55 samples, prostate adenocarcinoma (PRAD-US): 52 samples, and liver hepatocellular carcinoma (LIHC-US): 48 samples. This data combined with the gene down-regulation data from breast cancer could be very useful to evaluate how different cancer types (based on tissue) are similar (based on genetic profiles). If we are able to classify tumors based on genetic profiles and exploit

current therapies that are based on cancer type, there is the potential to repurpose several clinically relevant chemotherapeutics for patient specific therapeutics.

7.1.3 The 1,745 genes down-regulated in breast cancer

Analysis of the TCGA data generated a list of 1,745 genes that were down-regulated in breast cancer, of which only 15 were identified in HYGIN and used as the basis for the breast cancer subnetwork. The remaining 1,730 genes that were identified have not been analyzed. Potential further analyses include comparing these genes to the SynLethDB to see if any of these genes have been identified to have SL partners in other studies or if other cancer types are analyzed, comparing the set of down-regulated genes across cancer types in order to identify common genes for novel drug development. They could also be used as a starting point for SL analysis *in vitro*. Genes of interest could be screened against the entire human genome in order to identify the complete set of SL interactions for any given gene that is down-regulated in breast cancer. The results would provide valid drug targets that can be exploited for patient therapeutics.

7.1.4 Additional FBXW7 analyses

The analyses surrounding *FBXW7* in this thesis are mostly computational. Although the drug data in the cancerRXgene database is from previous cell line experimentation, no wet lab experimentation was completed directly for the purpose of this thesis. There has been previous research that has identified SL interactions with *FBXW7 in vitro* [106], although there was no overlap between the 32 SL interactions identified in this thesis and that study. As a result, the natural progression for this research is to test *FBXW7* and its SL interactions *in vitro* and *in vivo* to provide biological evidence for these interactions.

7.2 Conclusion

Yeast genetic interaction data and orthology mapping was used to generate HYGIN, a humanized SL interaction network, containing 10,419 proposed SL interactions between 1,009

human genes. HYGIN improves on previous attempts to identify human SL interactions using yeast data by using more stringent and rigorous analyses. Cancerous and normal tissue data from the TCGA were evaluated to identify genes that were down-regulated in breast cancer. The genes that were in HYGIN and down-regulated in breast cancer were used to generate a breast cancer specific subnetwork containing 130 genes and 130 proposed SL interactions. The tumor suppressor FBXW7 was identified to have the most SL interactions in the breast cancer subnetwork and was selected for further study. *In silico* cell line and drug analysis was used to identify potential drugs for therapeutics when FBXW7 is down-regulated. Given that our hypothesis was to identify gene targets for breast cancer therapeutics, identifying 32 potential novel drug targets for FBXW7 deficient breast cancers makes this project a success. These findings provide the basis for further testing of SL interactions with FBXW7 *in vitro* and *in vivo* to validate them for potential patient therapeutics.

References

[1] Shahab Akhoondi, Dahui Sun, Natalie von der Lehr, Sophia Apostolidou, Kathleen Klotz, Alena Maljukova, Diana Cepeda, Heidi Fiegl, Dimitra Dafou, Christian Marth, Elisabeth Mueller-Holzner, Martin Corcoran, Markus Dagnell, Sepideh Zabihi Nejad, Babak Noori Nayer, Mohammad Reza Zali, Johan Hansson, Susanne Egyhazi, Fredrik Petersson, Per Sangfelt, Hans Nordgren, Dan Grander, Steven I Reed, Martin Widschwendter, Olle Sangfelt, and Charles Spruck. Fbxw7/hcdc4 is a general tumor suppressor in human cancer. *Cancer Res*, 67(19):9006–9012, 2007.

[2] Rolf Apweiler, Amos Bairoch, Cathy H. Wu, Winona C. Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, Maria J. Martin, Darren A. Natale, Claire O'Donovan, Nicole Redaschi, and Lai-Su L. Yeh. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research*, 32(Database issue):D115–119, 2004.

[3] Ramy K Aziz, Jonathan M Monk, Robert M Lewis, Suh In Loh, Arti Mishra, Amrita Abhay Nagle, Chitkala Satyanarayana, Saravanakumar Dhakshinamoorthy, Michele Luche, Douglas B Kitchen, Kathleen A Andrews, Nicole L Fong, Howard J Li, Bernhard O Palsson, and Pep Charusanti. Systems biology-guided identification of synthetic lethal gene pairs and its potential use to discover antibiotic combinations. *Scientific Reports*, 5:16025, 2015.

[4] Melanie L. Bailey, Tejomayee Singh, Patricia Mero, Jason Moffat, and Philip Hieter. Dependence of Human Colorectal Cells Lacking the FBW7 Tumor Suppressor on the Spindle Assembly Checkpoint. *Genetics*, 201(3):885–895, 2015.

[5] S J Baker, E R Fearon, J M Nigro, S R Hamilton, A C Preisinger, J M Jessup, P vanTuinen, D H Ledbetter, D F Barker, Y Nakamura, R White, and B Vogelstein. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science*, 244(4901):217–221, 1989.

[6] Adriano Barbosa-Silva, Jean-Fred Fontaine, Elisa R Donnard, Fernanda Stussi, J Miguel Ortega, and Miguel A Andrade-Navarro. Pescador, a web-based tool to assist text-mining of biointeractions extracted from pubmed queries. *BMC Bioinformatics*, 12:435, 2011.

[7] Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, Adam A. Margolin, Sungjoon Kim, Christopher J. Wilson, Joseph Lehár, Gregory V. Kryukov, Dmitriy Sonkin, Anupama Reddy, Manway Liu, Lauren Murray, Michael F. Berger, John E. Monahan, Paula Morais, Jodi Meltzer, Adam Korejwa, Judit Jané-Valbuena,

Felipa A. Mapa, Joseph Thibault, Eva Bric-Furlong, Pichai Raman, Aaron Shipway, Ingo H. Engels, Jill Cheng, Guoying K. Yu, Jianjun Yu, Peter Aspesi, Melanie de Silva, Kalpana Jagtap, Michael D. Jones, Li Wang, Charles Hatton, Emanuele Palescandolo, Supriya Gupta, Scott Mahan, Carrie Sougnez, Robert C. Onofrio, Ted Liefeld, Laura MacConaill, Wendy Winckler, Michael Reich, Nanxin Li, Jill P. Mesirov, Stacey B. Gabriel, Gad Getz, Kristin Ardlie, Vivien Chan, Vic E. Myer, Barbara L. Weber, Jeff Porter, Markus Warmuth, Peter Finan, Jennifer L. Harris, Matthew Meyerson, Todd R. Golub, Michael P. Morrissey, William R. Sellers, Robert Schlegel, and Levi A. Garraway. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483:603 EP –, 2012.

[8] Jirina Bartkova, Zuzana Hořejší, Karen Koed, Alwin Krämer, Frederic Tort, Karsten Zieger, Per Guldberg, Maxwell Sehested, Jahn M. Nesland, Claudia Lukas, Torben Ørntoft, Jiri Lukas, and Jiri Bartek. Dna damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*, 434:864 EP –, 2005.

[9] Stephen B. Baylin and Joyce E. Ohm. Epigenetic gene silencing in cancer –a mechanism for early oncogenic pathway addiction? *Nature Reviews Cancer*, 6:107 EP –, 2006.

[10] René Bernards. A Missing Link in Genotype-Directed Cancer Therapy. *Cell*, 151(3):465–468, 2012.

[11] Mary-Ann Bjornsti. Cancer therapeutics in yeast. *Cancer Cell*, 2(4):267 – 273, 2002.

[12] Joan Boren, Antonio Ramos Montoya, Pedro de Atauri, Begoña Comin-Anduix, Antonio Cortes, Josep J. Centelles, Wilma M. Frederiks, Cornelis J. F. Van Noorden, and Marta Cascante. Metabolic Control Analysis Aimed at the Ribose Synthesis Pathways of Tumor Cells: A New Strategy for Antitumor Drug Development. *Molecular Biology Reports*, 29(1-2):7–12, 2002.

[13] Laszlo G. Boros, Joaquim Puigjaner, Marta Cascante, Wai-Nang Paul Lee, James L. Brandes, Sara Bassilian, Fouza I. Yusuf, Robert D. Williams, Pete Muscarella, W. Scott Melvin, and William J. Schirmer. Oxythiamine and Dehydroepiandrosterone Inhibit the Nonoxidative Synthesis of Ribose and Tumor Cell Proliferation. *Cancer Research*, 57(19):4242–4248, 1997.

[14] James D Brenton, Lisa A Carey, Ahmed Ashour Ahmed, and Carlos Caldas. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J Clin Oncol*, 23(29):7350–7360, 2005.

[15] Calvin B. Bridges. The origin of variations in sexual and sex-limited characters. *American Naturalist*, 56(51–63), 1922.

[16] Matthew D Burstein, Anna Tsimelzon, Graham M Poage, Kyle R Covington, Alejandro Contreras, Suzanne A W Fuqua, Michelle I Savage, C Kent Osborne, Susan G Hilsenbeck, Jenny C Chang, Gordon B Mills, Ching C Lau, and Powel H Brown. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res*, 21(7):1688–1698, 2015.

[17] Eric S. Calhoun, Jessa B. Jones, Raheela Ashfaq, Volkan Adsay, Suzanne J. Baker, Virginia Valentine, Paula M. Hempen, Werner Hilgers, Charles J. Yeo, Ralph H. Hruban, and Scott E. Kern. Braf and fbxw7 (cdc4, fbw7, ago, sel10) mutations in distinct subsets of pancreatic cancer: Potential therapeutic targets. *The American Journal of Pathology*, 163(4):1255 – 1260, 2003.

[18] Jun Cao, Ming-Hua Ge, and Zhi-Qiang Ling. Fbxw7 Tumor Suppressor: A Vital Regulator Contributes to Human Tumorigenesis. *Medicine*, 95(7), 2016.

[19] Marta Cascante, Laszlo G. Boros, Begoña Comin-Anduix, Pedro de Atauri, Josep J. Centelles, and Paul W.-N. Lee. Metabolic control analysis in drug discovery and disease. *Nature Biotechnology*, 20(3):243–249, 2002.

[20] B. B. Chen, J. R. Glasser, T. A. Coon, and R. K. Mallampalli. Skp-cullin-F box E3 ligase component FBXL2 ubiquitinates Aurora B to inhibit tumorigenesis. *Cell Death & Disease*, 4(8):e759, 2013.

[21] Begoña Comín-Anduix, Joan Boren, Sonia Martinez, Cristina Moro, Josep J. Centelles, Raisa Trebukhina, Nataly Petushok, Wai-Nang Paul Lee, Laszlo G. Boros, and Marta Cascante. The effect of thiamine supplementation on tumour proliferation. *European Journal of Biochemistry*, 268(15):4177–4182, 2001.

[22] Michael Costanzo, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D. Spear, Carolyn S. Sevier, Huiming Ding, Judice L.Y. Koh, Kiana Toufighi, Sara Mostafavi, Jeany Prinz, Robert P. St. Onge, Benjamin VanderSluis, Taras Makhnevych, Franco J. Vizeacoumar, Solmaz Alizadeh, Sondra Bahr, Renee L. Brost, Yiqun Chen, Murat Cokol, Raamesh Deshpande, Zhijian Li, Zhen-Yuan Lin, Wendy Liang, Michaela Marback, Jadine Paw, Bryan-Joseph San Luis, Ermira Shuteriqi, Amy Hin Yan Tong, Nydia van Dyk, Iain M. Wallace, Joseph A. Whitney, Matthew T. Weirauch, Guoqing Zhong, Hongwei Zhu, Walid A. Houry, Michael Brudno, Sasan Ragibizadeh, Balázs Papp, Csaba Pál, Frederick P. Roth, Guri Giaever, Corey Nislow, Olga G. Troyanskaya, Howard Bussey, Gary D. Bader, Anne-Claude Gingras, Quaid D. Morris, Philip M. Kim, Chris A. Kaiser, Chad L. Myers, Brenda J. Andrews, and Charles Boone. The Genetic Landscape of a Cell. *Science*, 327(5964):425–431, 2010.

[23] Michael Costanzo, Benjamin VanderSluis, Elizabeth N. Koch, Anastasia Baryshnikova, Carles Pons, Guihong Tan, Wen Wang, Matej Usaj, Julia Hanchard, Susan D. Lee, Vicent Pelechano, Erin B. Styles, Maximilian Billmann, Jolanda van Leeuwen, Nydia van Dyk, Zhen-Yuan Lin, Elena Kuzmin, Justin Nelson, Jeff S. Piotrowski, Tharan Srikumar, Sondra Bahr, Yiqun Chen, Raamesh Deshpande, Christoph F. Kurat, Sheena C. Li, Zhijian Li, Mojca Mattiazzi Usaj, Hiroki Okada, Natasha Pascoe, Bryan-Joseph San Luis, Sara Sharifpoor, Emira Shuteriqi, Scott W. Simpkins, Jamie Snider, Harsha Garadi Suresh, Yizhao Tan, Hongwei Zhu, Noel Malod-Dognin, Vuk Janjic, Natasa Przulj, Olga G. Troyanskaya, Igor Stagljar, Tian Xia, Yoshikazu Ohya, Anne-Claude Gingras, Brian Raught, Michael Boutros, Lars M. Steinmetz, Claire L. Moore, Adam P. Rosebrock, Amy A. Caudy, Chad L. Myers, Brenda Andrews, and Charles

Boone. A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 353(6306):aaf1420, 2016.

[24] Heim D., Budczies J., Stenzinger A., Treue D., Hufnagl P., Denkert C., Dietel M., and Klauschen F. Cancer beyond organ and tissue specificity: Next-generation-sequencing gene mutation data reveal complex genetic similarities across major cancers. *International Journal of Cancer*, 135(10):2362–2369, 2014.

[25] Chi V Dang. Myc on the path to cancer. *Cell*, 149(1):22–35, 2012.

[26] Ryan J. Davis, Markus Welcker, and Bruce E. Clurman. Tumor Suppression by the FBW7 Ubiquitin Ligase: Mechanisms and Opportunities. *Cancer Cell*, 26(4):455–464, 2014.

[27] Raamesh Deshpande, Michael Asiedu, Mitchell Klebig, Shari Sutor, Elena Kuzmin, Justin Nelson, Jeff Piotrowski, Seung Ho Shin, Minoru Yoshida, Michael Costanzo, Charles Boone, Dennis A. Wigle, and Chad L. Myers. A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Research*, 73(20):6128–36, 2013.

[28] Jay L. Devore and Kenneth N. Berk. *Modern Mathematical Statistics with Applications*. Springer Texts in Statistics. Springer, 2nd edition, 2012.

[29] S. J. Dixon, Y. Fedyshyn, J. L. Y. Koh, T. S. K. Prasad, C. Chahwan, G. Chua, K. Toufighi, A. Baryshnikova, J. Hayles, K.-L. Hoe, D.-U. Kim, H.-O. Park, C. L. Myers, A. Pandey, D. Durocher, B. J. Andrews, and C. Boone. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. *Proceedings of the National Academy of Sciences*, 105(43):16653–16658, 2008.

[30] Scott J Dixon, Brenda J Andrews, and Charles Boone. Exploring the conservation of synthetic lethal genetic interaction networks. *Communicative & Integrative Biology*, 2(2):78–81, 2009.

[31] L A Donehower, M Harvey, B L Slagle, M J McArthur, C A Jr Montgomery, J S Butel, and A Bradley. Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature*, 356(6366):215–221, 1992.

[32] Yulia Einav, Reuven Agami, and Dan Canaani. shRNA-mediated RNA interference as a tool for genetic synthetic lethality screening in mouse embryo fibroblasts. *FEBS Letters*, 579(1):199–202, 2005.

[33] Manel Esteller. Cancer epigenomics: Dna methylomes and histone-modification maps. *Nature Reviews Genetics*, 8:286 EP –, 2007.

[34] Elisa Ferrari, Chiara Lucca, and Marco Foiani. A lethal combination for cancer cells: Synthetic lethality screenings for drug discovery. *European Journal of Cancer*, 46(16):2889–2895, 2010.

[35] Joanne S. Finch, Margaret E. Tome, Kevin A. Kwei, and G. Tim Bowden. Catalase reverses tumorigenicity in a malignant cell line by an epidermal growth factor receptor pathway. *Free Radical Biology and Medicine*, 40(5):863–875, 2006.

[36] Helen V. Firth, Shola M. Richards, A. Paul Bevan, Stephen Clayton, Manuel Corpas, Diana Rajan, Steven Van Vooren, Yves Moreau, Roger M. Pettett, and Nigel P. Carter. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *The American Journal of Human Genetics*, 84(4):524–533, 2009.

[37] Jean-Fred Fontaine, Adriano Barbosa-Silva, Martin Schaefer, Matthew R Huska, Enrique M Muro, and Miguel A Andrade-Navarro. Medlineranker: flexible ranking of biomedical literature. *Nucleic Acids Res*, 37(Web Server issue):W141–6, 2009.

[38] Carlos M Galmarini, John R Mackey, and Charles Dumontet. Nucleoside analogues and nucleobases in cancer treatment. *The Lancet Oncology*, 3(7):415–424, 2002.

[39] Mercedes Garcia-Gil, Maria Grazia Tozzi, Stefano Varani, Lorenza Della Verde, Edoardo Petrotto, Francesco Balestri, Laura Colombaioni, and Marcella Camici. The combination of adenosine deaminase inhibition and deoxyadenosine induces apoptosis in a human astrocytoma cell line. *Neurochemistry International*, 80:14–22, 2015.

[40] Kumar Gaurav and Yasha Hasija. Comparative analysis of metabolic network of pathogens. *Frontiers in Biology*, 12(2):139–150, 2017.

[41] Guri Giaever, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, Ankuta Lucau-Danila, Keith Anderson, Bruno André, Adam P. Arkin, Anna Astromoff, Mohamed El Bakkoury, Rhonda Bangham, Rocio Benito, Sophie Brachat, Stefano Campanaro, Matt Curtiss, Karen Davis, Adam Deutschbauer, Karl-Dieter Entian, Patrick Flaherty, Francoise Foury, David J. Garfinkel, Mark Gerstein, Deanna Gotte, Ulrich Güldener, Johannes H. Hegemann, Svenja Hempel, Zelek Herman, Daniel F. Jaramillo, Diane E. Kelly, Steven L. Kelly, Peter Kötter, Darlene LaBonte, David C. Lamb, Ning Lan, Hong Liang, Hong Liao, Lucy Liu, Chuanyun Luo, Marc Lussier, Rong Mao, Patrice Menard, Siew Loon Ooi, Jose L. Revuelta, Christopher J. Roberts, Matthias Rose, Petra Ross-Macdonald, Bart Scherens, Greg Schimmack, Brenda Shafer, Daniel D. Shoemaker, Sharon Sookhai-Mahadeo, Reginald K. Storms, Jeffrey N. Strathern, Giorgio Valle, Marleen Voet, Guido Volckaert, Ching-yun Wang, Teresa R. Ward, Julie Wilhelmy, Elizabeth A. Winzeler, Yonghong Yang, Grace Yen, Elaine Youngman, Kexin Yu, Howard Bussey, Jef D. Boeke, Michael Snyder, Peter Philippsen, Ronald W. Davis, and Mark Johnston. Functional profiling of the saccharomyces cerevisiae genome. *Nature*, 418:387 EP –, 2002.

[42] Clemens Grabher, Harald von Boehmer, and A Thomas Look. Notch 1 activation in the molecular pathogenesis of t-cell acute lymphoblastic leukaemia. *Nat Rev Cancer*, 6(5):347–359, 2006.

[43] Nicoletta Guaragnella, Vanessa Palermo, Alvaro Galli, Loredana Moro, Cristina Mazzoni, and Sergio Giannattasio. The expanding role of yeast in cancer research and

diagnosis: insights into the function of the oncosuppressors p53 and brca1/2. *FEMS Yeast Res*, 14(1):2–16, 2014.

[44] Oriol Güell, Francesc Sagués, and M. Ángeles Serrano. Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLOS Computational Biology*, 10(5):1–10, 2014.

[45] Jing Guo, Hui Liu, and Jie Zheng. SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Research*, 44(Database issue):D1011–D1017, 2016.

[46] Bing Hao, Stephanie Oehlmann, Mathew E Sowa, J Wade Harper, and Nikola P Pavletich. Structure of a fbw7-skp1-cyclin e complex: multisite-phosphorylated substrate recognition by scf ubiquitin ligases. *Mol Cell*, 26(1):131–143, 2007.

[47] Richard Harrison, Balázs Papp, Csaba Pál, Stephen G Oliver, and Daniela Delneri. Plasticity of genetic interactions in metabolic networks of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 104(7):2307–2312, 2007.

[48] Traver Hart, Megha Chandrashekhar, Michael Aregger, Zachary Steinhart, Kevin R. Brown, Graham MacLeod, Monika Mis, Michal Zimmermann, Amelie Fradet-Turcotte, Song Sun, Patricia Mero, Peter Dirks, Sachdev Sidhu, Frederick P. Roth, Olivia S. Rissland, Daniel Durocher, Stephane Angers, and Jason Moffat. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, 163(6):1515–1526, 2015.

[49] L H Hartwell. Macromolecule synthesis in temperature-sensitive mutants of yeast. *J Bacteriol*, 93(5):1662–1670, 1967.

[50] Leland H. Hartwell, Joseph Culotti, and Brian Reid. Genetic control of the cell-division cycle in yeast, i. detection of mutants. *Proceedings of the National Academy of Sciences*, 66(2):352–359, 1970.

[51] John I. Hendry, Charulata B. Prasannan, Aditi Joshi, Santanu Dasgupta, and Pramod P. Wangikar. Metabolic model of synechococcus sp. pcc 7002: Prediction of flux distribution and network modification for enhanced biofuel production. *Bioresource Technology*, 213:190 – 197, 2016. International Conference on New Horizons in Biotechnology (NHBT-2015).

[52] Katherine A Hoadley, Christina Yau, Toshinori Hinoue, Denise M Wolf, Alexander J Lazar, Esther Drill, Ronglai Shen, Alison M Taylor, Andrew D Cherniack, Vesteinn Thorsson, Rehan Akbani, Reanne Bowlby, Christopher K Wong, Maciej Wiznerowicz, Francisco Sanchez-Vega, A Gordon Robertson, Barbara G Schneider, Michael S Lawrence, Houtan Noushmehr, Tathiane M Malta, Joshua M Stuart, Christopher C Benz, and Peter W Laird. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell*, 173(2):291–304, 2018.

[53] Katherine A. Hoadley, Christina Yau, Denise M. Wolf, Andrew D. Cherniack, David Tamborero, Sam Ng, Max D.M. Leiserson, Beifang Niu, Michael D. McLellan, Vladislav Uzunangelov, Jiashan Zhang, Cyriac Kandoth, Rehan Akbani, Hui Shen, Larsson Omberg, Andy Chu, Adam A. Margolin, Laura J. van't Veer, Nuria Lopez-Bigas, Peter W. Laird, Benjamin J. Raphael, Li Ding, A. Gordon Robertson, Lauren A. Byers, Gordon B. Mills, John N. Weinstein, Carter Van Waes, Zhong Chen, Eric A. Collisson, Christopher C. Benz, Charles M. Perou, and Joshua M. Stuart. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, 158(4):929 – 944, 2014.

[54] M Hollstein, D Sidransky, B Vogelstein, and C C Harris. p53 mutations in human cancers. *Science*, 253(5015):49–53, 1991.

[55] Jaime Huerta-Cepas, Salvador Capella-Gutierrez, Leszek P Pryszcz, Ivan Denisov, Diego Kormes, Marina Marcet-Houben, and Toni Gabaldon. Phylomedb v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res*, 39(Database issue):D556–60, 2011.

[56] Jaime Huerta-Cepas, Damian Szklarczyk, Kristoffer Forslund, Helen Cook, Davide Heller, Mathias C. Walter, Thomas Rattei, Daniel R. Mende, Shinichi Sunagawa, Michael Kuhn, Lars Juhl Jensen, Christian von Mering, and Peer Bork. eggnog 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research*, 44(D1):D286–D293, 2016.

[57] Meenakshi Jain, Constadina Arvanitis, Kenneth Chu, William Dewey, Edith Leonhardt, Maxine Trinh, Christopher D Sundberg, J Michael Bishop, and Dean W Felsher. Sustained loss of a neoplastic phenotype by brief inactivation of myc. *Science*, 297(5578):102–104, 2002.

[58] Livnat Jerby-Arnon, Nadja Pfetzer, Yedael Y. Waldman, Lynn McGarry, Daniel James, Emma Shanks, Brinton Seashore-Ludlow, Adam Weinstock, Tamar Geiger, Paul A. Clemons, Eyal Gottlieb, and Eytan Ruppin. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell*, 158(5):1199–1209, 2014.

[59] Pascal Jezequel, Delphine Loussouarn, Catherine Guerin-Charbonnel, Loic Campion, Antoine Vanier, Wilfried Gouraud, Hamza Lasla, Catherine Guette, Isabelle Valo, Veronique Verriele, and Mario Campone. Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response. *Breast Cancer Res*, 17:43, 2015.

[60] Peter A. Jones and Stephen B. Baylin. The fundamental role of epigenetic events in cancer. *Nature Reviews Genetics*, 3:415 EP –, 2002.

[61] Javed Khan, Jun S. Wei, Markus Ringnér, Lao H. Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R. Antonescu, Carsten Peterson, and Paul S. Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7:673 EP –, 2001.

[62] Takashi Kimura, Mitsukazu Gotoh, Yusuke Nakamura, and Hirofumi Arakawa. hcdc4b, a regulator of cyclin e, as a direct transcriptional target of p53. *Cancer Sci*, 94(5):431–436, 2003.

[63] Moon S Koh, Michael Ittmann, Dov Kadmon, Timothy C Thompson, and Fredrick S Leach. Cdc4 gene expression as potential biomarker for targeted therapy in prostate cancer. *Cancer Biol Ther*, 5(1):78–83, 2006.

[64] M H Kubbutat, R L Ludwig, A J Levine, and K H Vousden. Analysis of the degradation function of mdm2. *Cell Growth Differ*, 10(2):87–92, 1999.

[65] Nolwenn Le Meur and Robert Gentleman. Modeling synthetic lethality. *Genome Biology*, 9(9):R135–R135, 2008.

[66] Byung-Hoon Lee, Ying Lu, Miguel A. Prado, Yuan Shi, Geng Tian, Shuangwu Sun, Suzanne Elsasser, Steven P. Gygi, Randall W. King, and Daniel Finley. USP14 deubiquitinates proteasome-bound substrates that are ubiquitinated at multiple sites. *Nature*, 532(7599):398–401, 2016.

[67] Eva YHP Lee and William J Muller. Oncogenes and tumor suppressor genes. *Cold Spring Harbor Perspectives in Biology*, 2(10):a003236, 2010.

[68] Brian D. Lehmann, Joshua A. Bauer, Xi Chen, Melinda E. Sanders, A. Bapsi Chakravarthy, Yu Shyr, and Jennifer A. Pietenpol. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *The Journal of Clinical Investigation*, 121(7):2750–2767, 2011.

[69] Brian D. Lehmann, Bojana Jovanović, Xi Chen, Monica V. Estrada, Kimberly N. Johnson, Yu Shyr, Harold L. Moses, Melinda E. Sanders, and Jennifer A. Pietenpol. Refinement of triple-negative breast cancer molecular subtypes: Implications for neoadjuvant chemotherapy selection. *PLOS ONE*, 11(6):1–22, 2016.

[70] Li Li, Christian J Jr Stoeckert, and David S Roos. Orthomcl: identification of ortholog groups for eukaryotic genomes. *Genome Res*, 13(9):2178–2189, 2003.

[71] Jian-Hua Mao, Jesus Perez-losada, Di Wu, Reyno DelRosario, Ryosuke Tsunematsu, Keiichi I. Nakayama, Ken Brown, Sheila Bryson, and Allan Balmain. Fbxw7/cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature*, 432:775 EP –, 2004.

[72] Jian-Hua Mao, Jesus Perez-Losada, Di Wu, Reyno Delrosario, Ryosuke Tsunematsu, Keiichi I Nakayama, Ken Brown, Sheila Bryson, and Allan Balmain. Fbxw7/cdc4 is a p53-dependent, haploinsufficient tumour suppressor gene. *Nature*, 432(7018):775–779, 2004.

[73] Hiroko Masuda, Keith A. Baggerly, Ying Wang, Ya Zhang, Ana Maria Gonzalez-Angulo, Funda Meric-Bernstam, Vicente Valero, Brian D. Lehmann, Jennifer A. Pietenpol, Gabriel N. Hortobagyi, W. Fraser Symmans, and Naoto T. Ueno. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clinical Cancer Research*, 19(19):5533–5540, 2013.

[74] Dariusz Matlak and Ewa Szczurek. Epistasis in genomic and survival data of cancer patients. *PLOS Computational Biology*, 13(7):1–16, 2017.

[75] Renata Matuo, Fabricio G Sousa, Daniele G Soares, Diego Bonatto, Jenifer Saffi, Alexandre E Escargueil, Annette K Larsen, and Joao Antonio Pegas Henriques. Saccharomyces cerevisiae as a model system to study the response to anticancer agents. *Cancer Chemother Pharmacol*, 70(4):491–502, 2012.

[76] Kirk J. McManus, Irene J. Barrett, Yasaman Nouhi, and Philip Hieter. Specific synthetic lethal killing of RAD54b-deficient human colorectal cancer cells by FEN1 silencing. *Proceedings of the National Academy of Sciences*, 106(9):3276–3281, 2009.

[77] Vivien Measday, Kristin Baetz, Julie Guzzo, Karen Yuen, Teresa Kwok, Bilal Sheikh, Huiming Ding, Ryo Ueta, Trinh Hoac, Benjamin Cheng, Isabelle Pot, Amy Tong, Yuko Yamaguchi-Iwai, Charles Boone, Phil Hieter, and Brenda Andrews. Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13956–13961, 2005.

[78] Magali Michaut and Gary D. Bader. Multiple genetic interaction experiments provide complementary information useful for gene function prediction. *PLOS Computational Biology*, 8(6):1–12, 2012.

[79] L M Mulligan, G J Matlashewski, H J Scrable, and W K Cavenee. Mechanisms of p53 loss in human sarcomas. *Proceedings of the National Academy of Sciences*, 87(15):5863–5867, 1990.

[80] Arvindhan Nagarajan, Parmanand Malvi, and Narendra Wajapeyee. Oncogene-directed alterations in cancer cell metabolism. *Trends in cancer*, 2(7):365–377, 2016.

[81] Hao G Nguyen, Dharmaraj Chinnappan, Takeshi Urano, and Katya Ravid. Mechanism of aurora-b degradation and its dependency on intact ken and a-boxes: Identification of an aneuploidy-promoting property. *Molecular and Cellular Biology*, 25(12):4977–4992, 2005.

[82] J M Nigro, S J Baker, A C Preisinger, J M Jessup, R Hostetter, K Cleary, S H Bigner, N Davidson, S Baylin, and P Devilee. Mutations in the p53 gene occur in diverse human tumour types. *Nature*, 342(6250):705–708, 1989.

[83] N. Niitsu, Y. Yamaguchi, M. Umeda, and Y. Honma. Human monocytoid leukemia cells are highly sensitive to apoptosis induced by 2'-deoxycoformycin and 2'-deoxyadenosine: association with dATP-dependent activation of caspase-3. *Blood*, 92(9):3368–3375, 1998.

[84] Siew Loon Ooi, Xuewen Pan, Brian D. Peyser, Ping Ye, Pamela B. Meluh, Daniel S. Yuan, Rafael A. Irizarry, Joel S. Bader, Forrest A. Spencer, and Jef D. Boeke. Global synthetic-lethality analysis and yeast functional profiling. *Trends in Genetics*, 22(1):56–63, 2006.

[85] Breast Cancer Organization. Breast cancer information and support, November 2016.

[86] Stephen Orlicky, Xiaojing Tang, Andrew Willems, Mike Tyers, and Frank Sicheri. Structural basis for phosphodependent substrate selection and orientation by the scfcdc4 ubiquitin ligase. *Cell*, 112(2):243–256, 2003.

[87] Toshinori Ozaki and Akira Nakagawara. Role of p53 in cell death and human cancers. *Cancers*, 3(1):994–1013, 2011.

[88] Sri R Paladugu, Shan Zhao, Animesh Ray, and Alpan Raval. Mining protein networks for synthetic genetic interactions. *BMC Bioinformatics*, 9:426–426, 2008.

[89] James M. Paul, Shaina D. Templeton, Akanksha Baharani, Andrew Freywald, and Franco J. Vizeacoumar. Building high-resolution synthetic lethal networks: a 'Google map' of the cancer cell. *Trends in Molecular Medicine*, 20(12):704–715, 2014.

[90] James M. Paul, Behzad Toosi, Frederick S. Vizeacoumar, Kalpana Kalyanasundaram Bhanumathy, Yue Li, Courtney Gerger, Amr El Zawily, Tanya Freywald, Deborah H. Anderson, Darrell Mousseau, Rani Kanthan, Zhaolei Zhang, Franco J. Vizeacoumar, and Andrew Freywald. Targeting synthetic lethality between the SRC kinase and the EPHB6 receptor may benefit cancer treatment. *Oncotarget*, 7(31):50027–50042, 2016.

[91] Paola Perego, Gretchen S. Jimenez, Laura Gatti, Stephen B. Howell, and Franco Zunino. Yeast mutants as a model system for identification of determinants of chemosensitivity. *Pharmacological Reviews*, 52(4):477–492, 2000.

[92] Clara Pereira, Isabel Coutinho, Joana Soares, Claudia Bessa, Mariana Leao, and Lucilia Saraiva. New insights into cancer-related proteins provided by the yeast model. *FEBS J*, 279(5):697–712, 2012.

[93] C M Perou, T Sorlie, M B Eisen, M van de Rijn, S S Jeffrey, C A Rees, J R Pollack, D T Ross, H Johnsen, L A Akslen, O Fluge, A Pergamenschikov, C Williams, S X Zhu, P E Lonning, A L Borresen-Dale, P O Brown, and D Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–752, 2000.

[94] Aditya Pratapa, Shankar Balachandran, and Karthik Raman. Fast-SL: an efficient algorithm to identify synthetic lethal sets in metabolic networks. *Bioinformatics*, 31(20):3299–3305, 2015.

[95] Leslie Pray. L. H. Hartwell's Yeast: A Model Organism for Studying Somatic Mutations and Cancer. *Nature Education*, 1(1):183, 2008.

[96] Leszek P Pryszcz, Jaime Huerta-Cepas, and Toni Gabaldon. Metaphors: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Res*, 39(5):e32, 2011.

[97] Christopher D Putnam, Vincent Pennaneach, and Richard D Kolodner. Saccharomyces cerevisiae as a model system to define the chromosomal instability phenotype. *Mol Cell Biol*, 25(16):7226–7238, 2005.

[98] Badr Ras, Begoña Comin, Joaquim Puigjaner, James L Brandes, Edmond Creppy, Dominique Saboureau, Rachid Ennamany, Wai-Nang Paul Lee, Laszlo G Boros, and Marta Cascante. Oxythiamine and dehydroepiandrosterone induce a G1 phase cycle arrest in Ehrlich's tumor cells through inhibition of the pentose cycle. *FEBS Letters*, 456(1):113–118, 1999.

[99] Stéphane Rihet, Philippe Vielh, Jacques Camonis, Bruno Goud, Sylvie Chevillard, and Jean Gunzburg. Mutation status of genes encoding rhoa, rac1, and cdc42 gtpases in a panel of invasive human colorectal and breast tumors. *Journal of Cancer Research and Clinical Oncology*, 127(12):733–738, 2001.

[100] Jue Ruan, Heng Li, Zhongzhong Chen, Avril Coghlan, Lachlan James M Coin, Yiran Guo, Jean-Karim Hériché, Yafeng Hu, Karsten Kristiansen, Ruiqiang Li, Tao Liu, Alan Moses, Junjie Qin, Søren Vang, Albert J Vilella, Abel Ureta-Vidal, Lars Bolund, Jun Wang, Jun Wang, and Richard Durbin. Treefam: 2008 update. *Nucleic acids research*, 36(Database issue):D735—40, 2008.

[101] Babu V. Sajesh, Brent J. Guppy, and Kirk J. McManus. Synthetic Genetic Targeting of Genome Instability in Cancer. *Cancers*, 5(3):739–761, 2013.

[102] Ophir Shalem, Neville E Sanjana, and Feng Zhang. High-throughput functional genomics using CRISPR-Cas9. *Nature Reviews Genetics*, 16(5):299–311, 2015.

[103] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[104] Julian A Simon. Yeast as a model system for anticancer drug discovery. *Emerging Therapeutic Targets*, 5(2):177–195, 2001. PMID: 15992175.

[105] Subarna Sinha, Daniel Thomas, Steven Chan, Yang Gao, Diede Brunen, Damoun Torabi, Andreas Reinisch, David Hernandez, Andy Chan, Erinn B. Rankin, Rene Bernards, Ravindra Majeti, and David L. Dill. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. *Nature Communications*, 8:15580, 2017.

[106] Subarna Sinha, Daniel Thomas, Steven Chan, Yang Gao, Diede Brunen, Damoun Torabi, Andreas Reinisch, David Hernandez, Andy Chan, Erinn B. Rankin, Rene Bernards, Ravindra Majeti, and David L. Dill. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. *Nature Communications*, 8:15580, 2017.

[107] Canadian Cancer Society. Canadian cancer statistics 2017, 2017.

[108] Chunge Song, Ruojin Ma, Xiaoyu Yang, and Sulei Pang. The Deubiquitinating Enzyme USP14 Regulates Leukemic Chemotherapy Drugs-Induced Cell Apoptosis by Suppressing Ubiquitination of Aurora Kinase B. *Cellular Physiology and Biochemistry*, 42(3):965–973, 2017.

[109] Erik L.L. Sonnhammer and Gabriel Östlund. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Research*, 43(Database issue):D234–D239, 2015.

[110] T Sorlie, C M Perou, R Tibshirani, T Aas, S Geisler, H Johnsen, T Hastie, M B Eisen, M van de Rijn, S S Jeffrey, T Thorsen, H Quist, J C Matese, P O Brown, D Botstein, P E Lonning, and A L Borresen-Dale. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98(19):10869–10874, 2001.

[111] C H Spruck, K A Won, and S I Reed. Deregulated cyclin e induces chromosome instability. *Nature*, 401(6750):297–300, 1999.

[112] Charles H Spruck, Heimo Strohmaier, Olle Sangfelt, Hannes M Muller, Michael Hubalek, Elisabeth Muller-Holzner, Christian Marth, Martin Widschwendter, and Steven I Reed. hcdc4 gene mutations in endometrial cancer. *Cancer Res*, 62(16):4535–4539, 2002.

[113] Rohith Srivas, John Paul Shen, Chih Cheng Yang, Su Ming Sun, Jianfeng Li, Andrew M. Gross, James Jensen, Katherine Licon, Ana Bojorquez-Gomez, Kristin Klepper, Justin Huang, Daniel Pekin, Jia L. Xu, Huwate Yeerna, Vignesh Sivaganesh, Leonie Kollenstart, Haico van Attikum, Pedro Aza-Blanc, Robert W. Sobol, and Trey Ideker. A network of conserved synthetic lethal interactions for exploration of precision cancer therapy. *Molecular Cell*, 63(3):514–525, 2016.

[114] Rohith Srivas, John Paul Shen, Chih Cheng Yang, Su Ming Sun, Jianfeng Li, Andrew M Gross, James Jensen, Katherine Licon, Ana Bojorquez-Gomez, Kristin Klepper, Justin Huang, Daniel Pekin, Jia L Xu, Huwate Yeerna, Vignesh Sivaganesh, Leonie Kollenstart, Haico van Attikum, Pedro Aza-Blanc, Robert W Sobol, and Trey Ideker. A network of conserved synthetic lethal interactions for exploration of precision cancer therapy. *Molecular Cell*, 63(3):514–525, 2016.

[115] Wanpeng Sun and Jian Yang. Functional mechanisms for human tumor suppressors. *Journal of Cancer*, 1:136–140, 2010.

[116] T Takahashi, M M Nau, I Chiba, M J Birrer, R K Rosenberg, M Vinocour, M Levitt, H Pass, A F Gazdar, and J D Minna. p53: a frequent target for genetic abnormalities in lung cancer. *Science*, 246(4929):491–494, 1989.

[117] Chieh-Lin Teng, Yun-Chi Hsieh, Liem Phan, Jihyun Shin, Chris Gully, Guermarie Velazquez-Torres, Stephen Skerl, Sai-Ching J. Yeung, Shih-Lan Hsu, and Mong-Hong Lee. FBXW7 is involved in Aurora B degradation. *Cell Cycle*, 11(21):4059–4068, 2012.

[118] David Thissen, Lynne Steinberg, and Daniel Kuang. Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27(1):77–83, 2002.

[119] D Thomas, L Kuras, R Barbey, H Cherest, P L Blaiseau, and Y Surdin-Kerjan. Met30p, a yeast transcriptional inhibitor that responds to s-adenosylmethionine, is an essential protein with wd40 repeats. *Molecular and Cellular Biology*, 15(12):6526–6534, 1995.

[120] Luis Tobalina, Jon Pey, Alberto Rezola, and Francisco J Planes. Assessment of fba based gene essentiality analysis in cancer with a fast context-specific network reconstruction method. *PLoS ONE*, 11(5):e0154583, 2016.

[121] R Todd and D T Wong. Oncogenes. *Anticancer Res*, 19(6A):4729–4746, 1999.

[122] Amy Hin Yan Tong, Guillaume Lesage, Gary D. Bader, Huiming Ding, Hong Xu, Xiaofeng Xin, James Young, Gabriel F. Berriz, Renee L. Brost, Michael Chang, YiQun Chen, Xin Cheng, Gordon Chua, Helena Friesen, Debra S. Goldberg, Jennifer Haynes, Christine Humphries, Grace He, Shamiza Hussein, Lizhu Ke, Nevan Krogan, Zhijian Li, Joshua N. Levinson, Hong Lu, Patrice Ménard, Christella Munyana, Ainslie B. Parsons, Owen Ryan, Raffi Tonikian, Tania Roberts, Anne-Marie Sdicu, Jesse Shapiro, Bilal Sheikh, Bernhard Suter, Sharyl L. Wong, Lan V. Zhang, Hongwei Zhu, Christopher G. Burd, Sean Munro, Chris Sander, Jasper Rine, Jack Greenblatt, Matthias Peter, Anthony Bretscher, Graham Bell, Frederick P. Roth, Grant W. Brown, Brenda Andrews, Howard Bussey, and Charles Boone. Global mapping of the yeast genetic interaction network. *Science*, 303(5659):808–813, 2004.

[123] Matej Usaj, Yizhao Tan, Wen Wang, Benjamin VanderSluis, Albert Zou, Chad L Myers, Michael Costanzo, Brenda Andrews, and Charles Boone. Thecellmap.org: A web-accessible database for visualizing and mining the global yeast genetic interaction network. *G3 (Bethesda)*, 7(5):1539–1549, 2017.

[124] Mukesh Verma. Personalized medicine and cancer. *Journal of Personalized Medicine*, 2(1):1–14, 2012.

[125] Carolina Vicente-Dueñas, Isabel Romero-Camarero, Cesar Cobaleda, and Isidro Sánchez-García. Function of oncogenes in cancer development: a changing paradigm. *The EMBO Journal*, 32(11):1502–1513, 2013.

[126] Pedro Vizán, Susana Sánchez-Tena, Gema Alcarraz-Vizán, Marta Soler, Ramon Messeguer, M. Dolors Pujol, Wai-Nang Paul Lee, and Marta Cascante. Characterization of the metabolic changes underlying growth factor angiogenic activation: identification of new potential therapeutic targets. *Carcinogenesis*, 30(6):946–952, 2009.

[127] Franco J Vizeacoumar, Roland Arnold, Frederick S Vizeacoumar, Megha Chandrashekhar, Alla Buzina, Jordan T F Young, Julian H M Kwan, Azin Sayad, Patricia Mero, Steffen Lawo, Hiromasa Tanaka, Kevin R Brown, Anastasia Baryshnikova, Anthony B Mak, Yaroslav Fedyshyn, Yadong Wang, Glauber C Brito, Dahlia Kasimer, Taras Makhnevych, Troy Ketela, Alessandro Datti, Mohan Babu, Andrew Emili, Laurence Pelletier, Jeff Wrana, Zev Wainberg, Philip M Kim, Robert Rottapel, Catherine A O'Brien, Brenda Andrews, Charles Boone, and Jason Moffat. A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. *Molecular Systems Biology*, 9:696, 2013.

[128] Yifan Wang and Binhua P Zhou. Fbw7-aurora b-p53 feedback loop regulates mitosis and cell growth. *Cell Cycle*, 11(22):4113–4114, 2012.

[129] Markus Welcker and Bruce E Clurman. Fbw7/hcdc4 dimerization regulates its substrate interactions. *Cell Div*, 2:7, 2007.

[130] Markus Welcker and Bruce E. Clurman. FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. *Nature Reviews Cancer*, 8(2):83–93, 2008.

[131] Sarah J. Welsh and Garth Powis. *Personalized Medicine for Cancer*, pages 93–107. Springer Netherlands, Dordrecht, 2008.

[132] Ligang Wu and Joel G. Belasco. Let me count the ways: Mechanisms of gene regulation by mirnas and sirnas. *Molecular Cell*, 29(1):1 – 7, 2008.

[133] Min Wu, Xuejuan Li, Fan Zhang, Xiaoli Li, Chee-Keong Kwoh, and Jie Zheng. In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. *Cancer Informatics*, 13(Suppl 3):71–80, 2014.

[134] W. Xie, R. Duan, and S. Safe. Estrogen induces adenosine deaminase gene expression in MCF-7 human breast cancer cells: role of estrogen receptor-Sp1 interactions. *Endocrinology*, 140(1):219–227, 1999.

[135] Clinton Yam, Sendurai A. Mani, and Stacy L. Moulder. Targeting the molecular subtypes of triple negative breast cancer: Understanding the diversity to progress the field. *The Oncologist*, 2017.

[136] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A Smith, I Richard Thompson, Sridhar Ramaswamy, P Andrew Futreal, Daniel A Haber, Michael R Stratton, Cyril Benes, Ultan McDermott, and Mathew J Garnett. Genomics of drug sensitivity in cancer (gdsc): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic acids research*, 41(Database issue):D955—61, 2013.

[137] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A. Smith, I. Richard Thompson, Sridhar Ramaswamy, P. Andrew Futreal, Daniel A. Haber, Michael R. Stratton, Cyril Benes, Ultan McDermott, and Mathew J. Garnett. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research*, 41(D1):D955–D961, 2013.

[138] Ping Ye, Brian D Peyser, Xuewen Pan, Jef D Boeke, Forrest A Spencer, and Joel S Bader. Gene function prediction from congruent synthetic lethal interactions in yeast. *Molecular Systems Biology*, 1:2005.0026–2005.0026, 2005.

[139] Karen W Y Yuen, Cheryl D Warren, Ou Chen, Teresa Kwok, Phil Hieter, and Forrest A Spencer. Systematic genome instability screens in yeast and their potential relevance

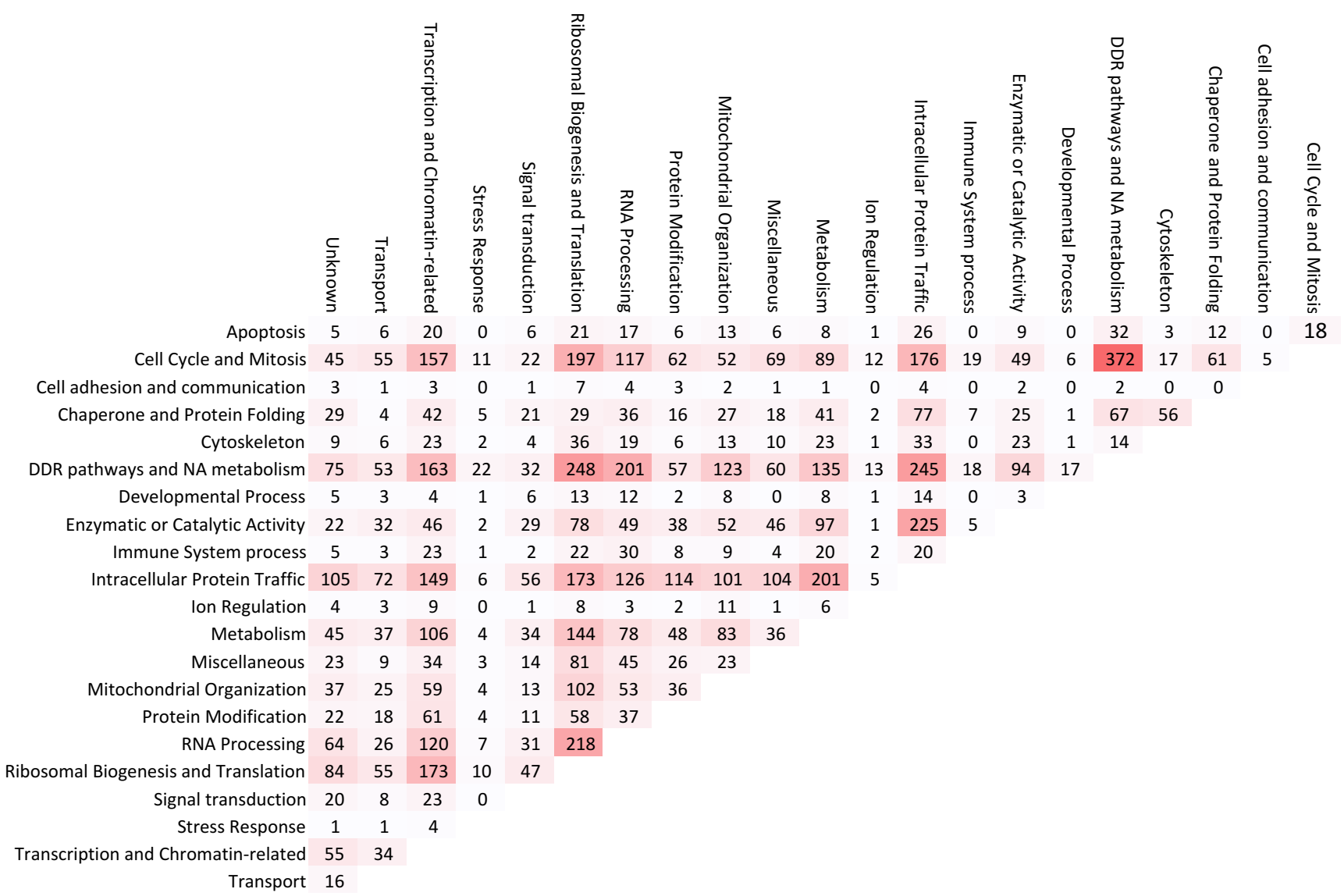
to cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 104(10):3925–3930, 2007.

[140] Barry R. Zeeberg, Joseph Riss, David W. Kane, Kimberly J. Bussey, Edward Uchio, W. Marston Linehan, J. Carl Barrett, and John N. Weinstein. Mistaken identifiers: Gene name errors can be introduced inadvertently when using excel in bioinformatics. *BMC Bioinformatics*, 5(1):80, 2004.

[141] Daniel R Zerbino, Premanand Achuthan, Wasiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, Laurent Gil, Leo Gordon, Leanne Haggerty, Erin Haskell, Thibaut Hourlier, Osagie G Izuogu, Sophie H Janacek, Thomas Juettemann, Jimmy Kiang To, Matthew R Laird, Ilias Lavidas, Zhicheng Liu, Jane E Loveland, Thomas Maurel, William McLaren, Benjamin Moore, Jonathan Mudge, Daniel N Murphy, Victoria Newman, Michael Nuhn, Denye Ogeh, Chuang Kee Ong, Anne Parker, Mateus Patricio, Harpreet Singh Riat, Helen Schuilenburg, Dan Sheppard, Helen Sparrow, Kieron Taylor, Anja Thormann, Alessandro Vullo, Brandon Walts, Amonida Zadissa, Adam Frankish, Sarah E Hunt, Myrto Kostadima, Nicholas Langridge, Fergal J Martin, Matthieu Muffato, Emily Perry, Magali Ruffier, Dan M Staines, Stephen J Trevanion, Bronwen L Aken, Fiona Cunningham, Andrew Yates, and Paul Flicek. Ensembl 2018. *Nucleic Acids Research*, 46(D1):D754–D761, 2018.

Appendix A

Supplemental Figures



	Unknown	Transport	Transcription and Chromatin-related	Stress Response	Signal transduction	Ribosomal Biogenesis and Translation	RNA Processing	Protein Modification	Mitochondrial Organization	Miscellaneous	Metabolism	Ion Regulation	Intracellular Protein Traffic	Immune System process	Enzymatic or Catalytic Activity	Developmental Process	DDR pathways and NA metabolism	Cytoskeleton	Chaperone and Protein Folding	Cell adhesion and communication	Cell Cycle and Mitosis
Apoptosis	5	6	20	0	6	21	17	6	13	6	8	1	26	0	9	0	32	3	12	0	18
Cell Cycle and Mitosis	45	55	157	11	22	197	117	62	52	69	89	12	176	19	49	6	372	17	61	5	
Cell adhesion and communication	3	1	3	0	1	7	4	3	2	1	1	0	4	0	2	0	2	0	0		
Chaperone and Protein Folding	29	4	42	5	21	29	36	16	27	18	41	2	77	7	25	1	67	56			
Cytoskeleton	9	6	23	2	4	36	19	6	13	10	23	1	33	0	23	1	14				
DDR pathways and NA metabolism	75	53	163	22	32	248	201	57	123	60	135	13	245	18	94	17					
Developmental Process	5	3	4	1	6	13	12	2	8	0	8	1	14	0	3						
Enzymatic or Catalytic Activity	22	32	46	2	29	78	49	38	52	46	97	1	225	5							
Immune System process	5	3	23	1	2	22	30	8	9	4	20	2	20								
Intracellular Protein Traffic	105	72	149	6	56	173	126	114	101	104	201	5									
Ion Regulation	4	3	9	0	1	8	3	2	11	1	6										
Metabolism	45	37	106	4	34	144	78	48	83	36											
Miscellaneous	23	9	34	3	14	81	45	26	23												
Mitochondrial Organization	37	25	59	4	13	102	53	36													
Protein Modification	22	18	61	4	11	58	37														
RNA Processing	64	26	120	7	31	218															
Ribosomal Biogenesis and Translation	84	55	173	10	47																
Signal transduction	20	8	23	0																	
Stress Response	1	1	4																		
Transcription and Chromatin-related	55	34																			
Transport	16																				

Figure A.1: 22 GO Slim terms and the interactions between them as translated from yeast are represented in this table. Cells are colour-coded according to the number of interactions between two groups of GO terms, where darker cells represent more interactions and lighter cells represent fewer interactions.

A

Searching the species
Homo sapiens
for the identifier **met30** excluding inparalogs scoring below 0.05

xml

Inparalog and Orthologs cluster for Saccharomyces cerevisiae and Homo sapiens

Cluster 817

Protein ID	Species	Score	Bootstrap	Description	Alternative ID
P39014	Saccharomyces cerevisiae	1	43%	F-box protein MET30	MET30_YEAST (UniProt)
Q969H0	Homo sapiens	1	99%	F-box/WD repeat-containing protein 7	FBXW7_HUMAN (UniProt)

B

Searching the species
Homo sapiens
for the identifier **cdc4** excluding inparalogs scoring below 0.05
More than one protein found:

Inparanoid	Gene ID	Protein ID	Species	Description	Alternative ID
Cluster	CDC4	A3LV96	Scheffersomyces stipitis	F box protein, for ubiquitin-dependent degradation	A3LV96_PICST (UniProt)
Cluster	CDC4	Q5A9A6	Candida albicans	Potential ubiquitin ligase F-box subunit	Q5A9A6_CANAL (UniProt)

Figure A.2: Two searches of the InParanoid database were conducted to verify that FBXW7 has a strict one-to-one mapping. A) The human genome was searched for the protein MET30 (the yeast protein that is orthologous to FBXW7 in humans). As expected, a single ortholog mapping was identified between *Saccharomyces cerevisiae* and *Homo sapiens*. B) The protein CDC4, a synonym for FBXW7, in humans was subsequently searched for in humans. Finding no ortholog in human, InParanoid reported "hits" in other organisms. Results showed that there is no entry for CDC4 in humans, which means that there is no additional FBXW7 ortholog in *Saccharomyces cerevisiae*.

Gene Name	Bladder urothelial carcinoma	Breast invasive carcinoma	Cholangiocarcinoma	Colon adenocarcinoma	Esophageal carcinoma	Glioblastoma multiforme	Head and neck squamous cell carcinoma	Kidney chromophobe	Kidney renal clear cell carcinoma	Kidney renal papillary cell carcinoma	Liver hepatocellular carcinoma	Lung adenocarcinoma	Lung squamous cell carcinoma	Pheochromocytoma and paraganglioma	Prostate adenocarcinoma	Rectum adenocarcinoma	Stomach adenocarcinoma	Thyroid carcinoma	Uterine corpus endometrial carcinoma
ABCG2	5.88E-05	2.48E-46	2.15E-06	2.06E-25	NS	NS	5.88E-04	4.82E-02	9.48E-07	7.70E-15	5.47E-12	4.36E-31	4.88E-24	2.22E-02	1.14E-11	2.28E-07	1.48E-03	7.29E-04	9.46E-20
ABHD6	3.12E-04	3.84E-39	3.22E-06	4.48E-22	6.29E-04	1.02E-03	NS	6.60E-06	NS	NS	6.36E-21	7.45E-30	6.47E-22	1.07E-02	2.57E-18	1.55E-05	1.24E-10	7.18E-11	2.16E-09
ADHFE1	8.84E-08	1.06E-54	2.15E-06	3.44E-25	1.15E-06	9.57E-04	3.12E-09	4.57E-06	4.22E-26	NS	5.55E-14	9.48E-14	2.31E-17	5.58E-03	NS	2.93E-07	3.84E-16	2.33E-14	6.87E-17
AIFM2	NS	1.47E-30	NS	9.61E-03	NS	NS	NS	3.22E-10	7.76E-05	2.87E-02	1.03E-12	6.83E-13	NS	1.10E-02	2.68E-15	NS	6.47E-03	1.00E-12	4.49E-12
AMT	5.01E-04	1.90E-34	2.15E-06	6.14E-15	1.74E-05	NS	NS	2.21E-07	6.21E-29	4.58E-08	4.05E-12	NS	4.26E-11	NS	4.87E-12	1.67E-02	3.73E-10	4.75E-02	2.00E-03
CAT	2.20E-05	1.16E-60	2.15E-06	8.34E-19	1.71E-04	1.17E-02	3.51E-07	NS	8.33E-34	1.28E-16	9.87E-25	1.29E-35	5.58E-32	1.66E-03	2.19E-09	6.83E-05	1.13E-08	2.62E-02	4.11E-07
FBXW7	2.37E-02	3.48E-57	4.50E-04	2.17E-02	4.87E-03	8.77E-05	1.56E-12	4.58E-05	4.88E-20	7.95E-06	1.09E-05	1.60E-02	4.53E-02	2.59E-03	3.69E-06	1.38E-02	3.09E-04	1.08E-12	8.50E-07
HAAO	6.77E-12	3.16E-40	2.15E-06	1.74E-03	NS	5.75E-03	1.08E-08	6.29E-10	9.83E-06	9.35E-03	3.65E-19	1.90E-09	1.50E-09	3.38E-02	1.21E-14	NS	1.94E-03	6.98E-04	2.63E-17
LARP6	8.70E-05	2.46E-53	3.56E-05	NS	1.88E-03	9.77E-05	6.33E-04	3.42E-06	3.86E-11	1.74E-02	3.96E-03	1.13E-07	3.13E-10	2.68E-03	8.67E-04	NS	2.78E-07	NS	5.28E-11
LDHD	NS	7.42E-31	2.15E-06	1.37E-25	4.23E-05	9.77E-05	3.86E-19	1.09E-07	6.10E-31	9.57E-04	6.55E-23	1.14E-05	9.66E-13	1.13E-02	NS	4.31E-07	2.96E-10	4.94E-17	NS
MGLL	4.90E-11	6.02E-46	2.05E-02	8.72E-24	4.79E-02	1.37E-03	1.33E-22	NS	2.57E-21	3.27E-02	2.38E-12	1.53E-15	2.31E-31	2.99E-02	3.74E-05	1.55E-05	9.14E-07	1.45E-02	5.21E-07
MSRA	7.13E-05	1.91E-49	2.15E-06	4.22E-14	5.80E-05	2.11E-02	2.15E-17	8.65E-09	7.74E-08	4.80E-09	1.68E-22	4.08E-19	1.34E-30	1.71E-02	6.22E-06	3.87E-04	2.55E-10	7.84E-12	3.88E-02
PEMT	5.19E-05	8.23E-20	2.15E-06	2.19E-15	1.92E-02	9.00E-04	2.92E-02	9.87E-04	2.86E-02	1.47E-06	3.51E-22	4.40E-02	NS	NS	4.59E-03	4.77E-05	NS	3.71E-15	2.11E-03
RGN	1.61E-09	4.13E-52	2.15E-06	2.35E-06	7.82E-06	NS	2.07E-14	7.61E-03	4.79E-19	8.57E-10	3.80E-19	4.48E-17	1.02E-30	1.61E-03	4.03E-15	1.57E-03	3.31E-10	4.50E-11	1.32E-16
UGP2	2.17E-05	1.36E-46	2.15E-06	2.41E-26	1.14E-04	1.58E-02	6.27E-16	2.62E-03	4.67E-15	6.75E-05	3.98E-20	4.26E-05	NS	1.50E-03	1.49E-09	1.31E-07	4.34E-08	5.93E-08	1.22E-05

Figure A.3: The 15 genes that are the basis for the breast cancer subnetwork were analyzed across 24 different cancer types. The 19 cancer types where the majority of the 15 genes are down-regulated are depicted here. Significant values in blue denote statistically significant lower expression in pooled cancer tissue when compared to pooled normal tissue, values in red denote statistically higher expression in cancer, and NS values denote a non-significant change in gene expression between cancer and normal tissues. All P-values were calculated using the Mann-Whitney U test.

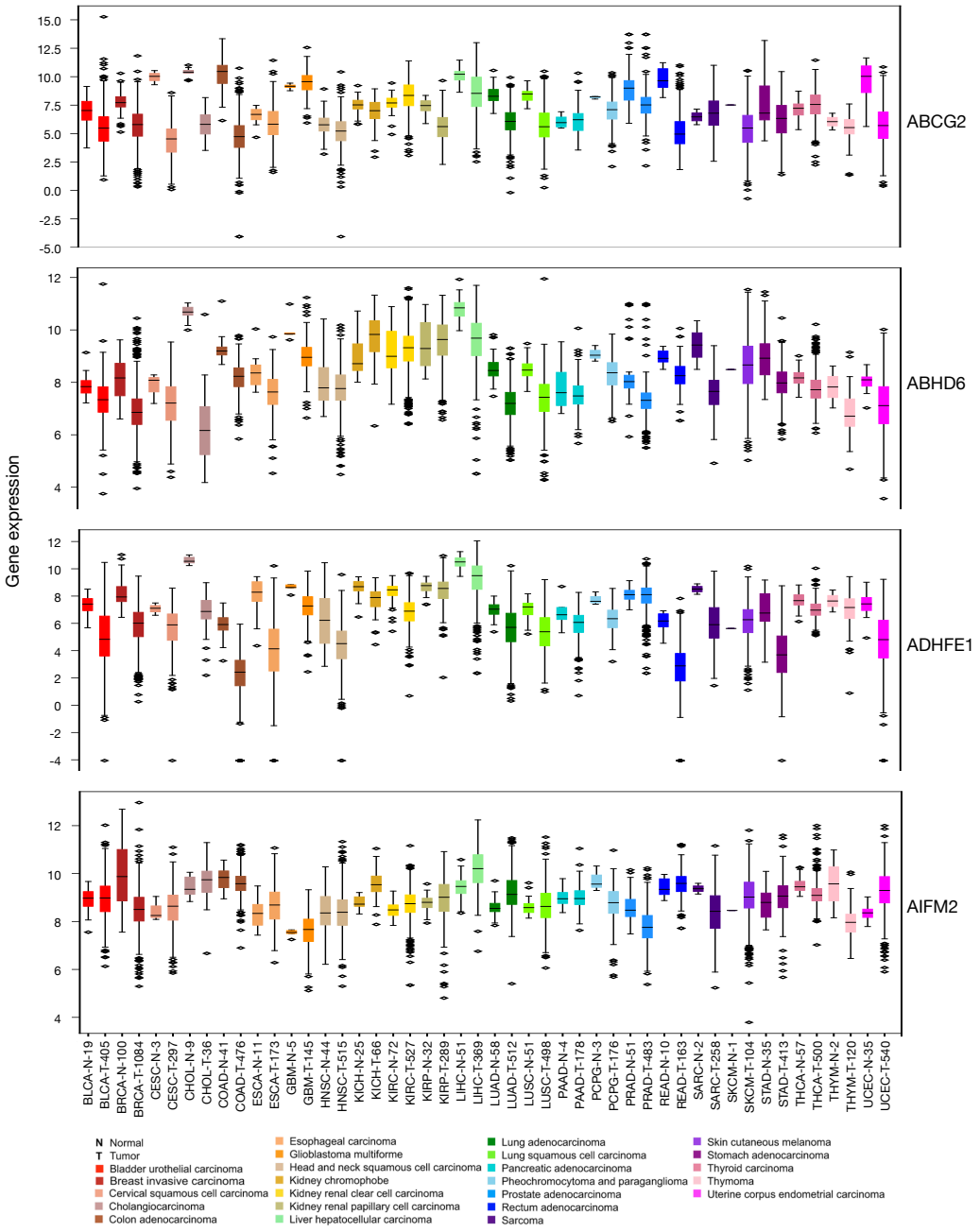


Figure A.4: Gene expression calculated by RNA-seq by Expectation-Maximization (log2) for *ABCG2*, *ABHD6*, *ADHFE1*, and *AIFM2* for 24 different cancer types comparing cancer and normal samples within the same tissue type. Box plots are colour-coded in pairs based on cancer type where the boxplot on the left is gene expression in cancer and the plot on the right is gene expression in normal tissue. Labels on the x-axis represent the abbreviation of the cancer type, whether the sample represents tumor (T) or normal (N) tissue, and the sample size.

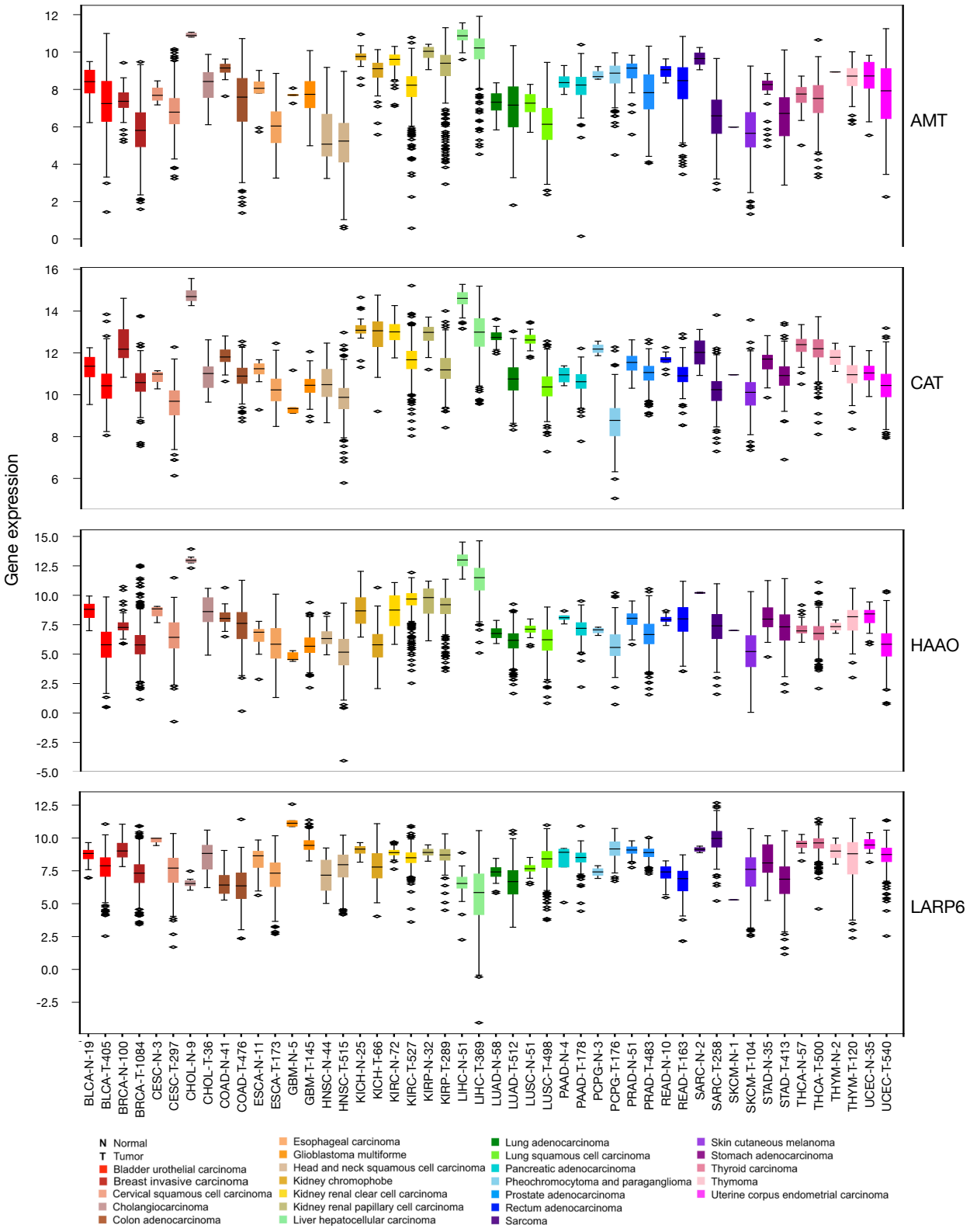


Figure A.5: Same as description for Figure A.4, except for *AMT*, *CAT*, *HAAO*, and *LARP6*.

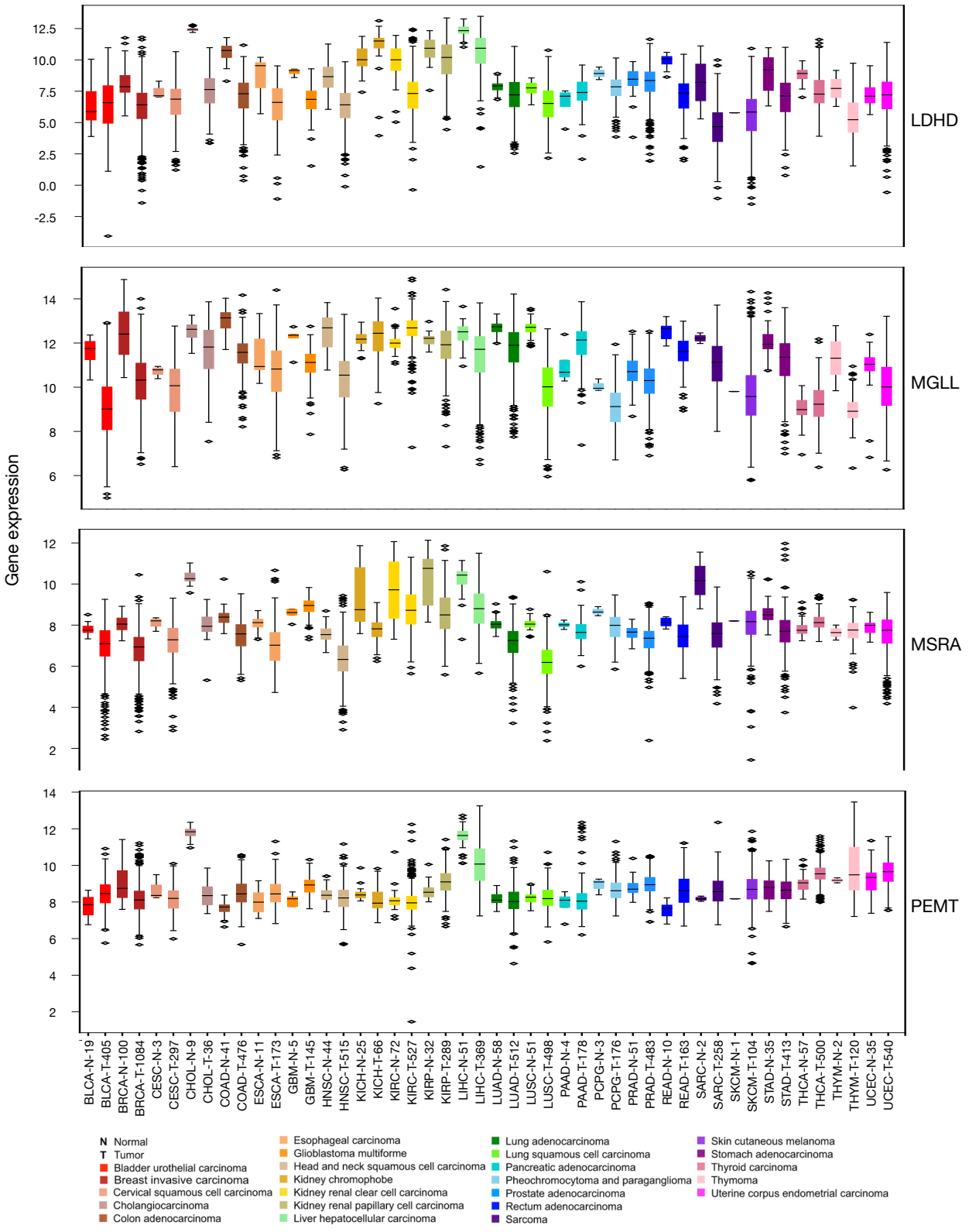


Figure A.6: Same as description for Figure A.4, except for *LDHD*, *MGLL*, *MSRA*, and *PEMT*.

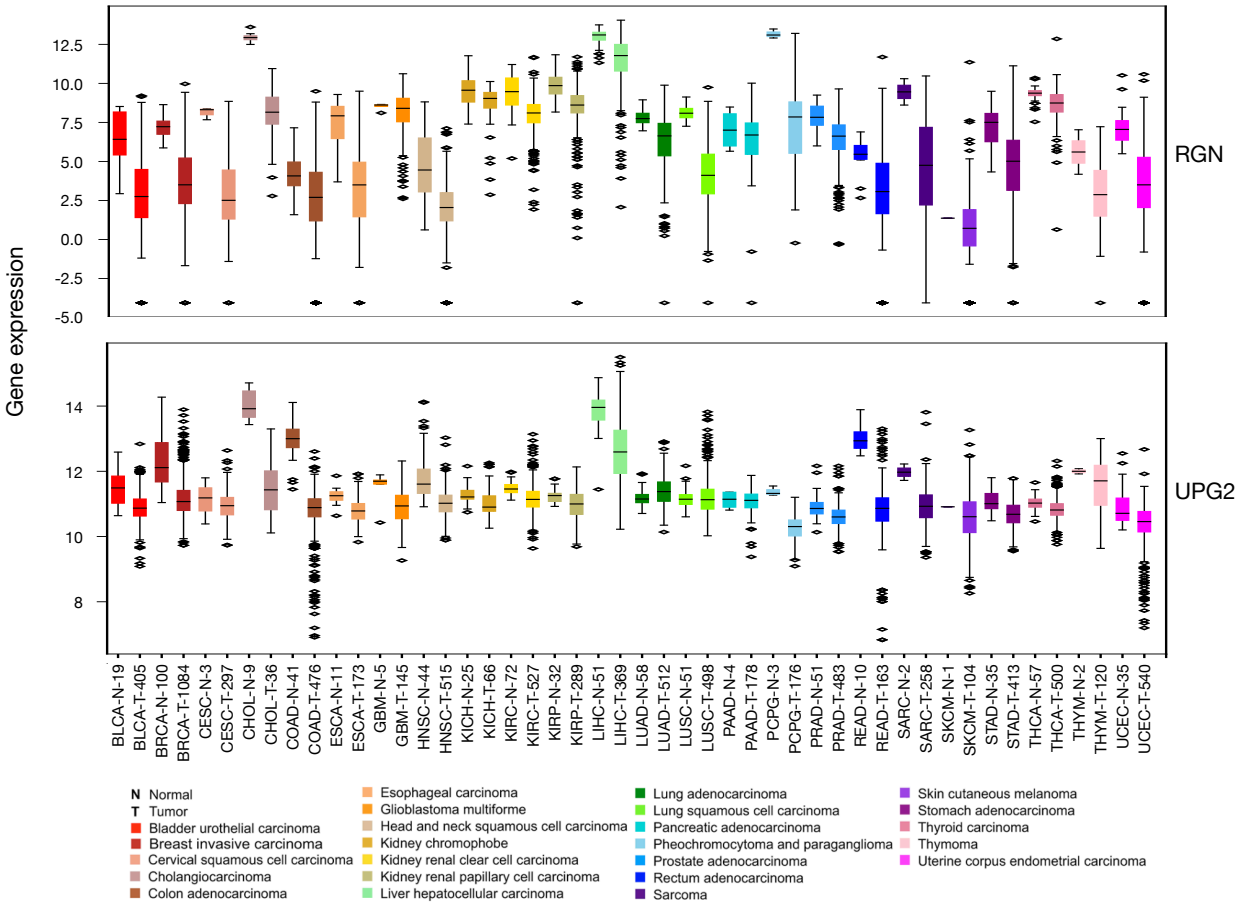


Figure A.7: Same as description for Figure A.4, except for *RGN* and *UPG2*.

Appendix B

Supplementary Tables

Supplementary tables included as comma separated value (csv) files. Table legends for each are provided below.

Supplementary Table B.1: Strict one-to-one ortholog mapping for yeast genes in humans. Ortholog mapping information was taken from InParanoid and filtered to identify instances where one yeast protein mapped to one human protein. The resulting 1,266 pairs of orthologs can be found in this table.

Supplementary Table B.2: HYGIN network table. This table contains the initial mapping data from the yeast network where SL interactions are represented in pairs by A and B (Yeast Gene ID A, Yeast Gene ID B, Yeast Gene Name A, Yeast Gene Name B, eValue, and pValue). Through the 3-stage mapping process used to generate HYGIN, it also contains columns corresponding to Yeast Protein Accession (A and B), Human Protein Accession (A and B), and Human Gene Name (A and B). Human gene names were used to generate the network in Cytoscape. Due to some reciprocal screens in the initial yeast screen, there were some duplicate (reciprocal) entries in this table. Once the network was loaded into Cytoscape,"Remove duplicate edges" was used to eliminate reciprocal hits. This resulted in the 1,009 genes and 10,419 edges in the final HYGIN network.

Supplementary Table B.3: GO Slim mapping table. Twenty-two GO Slims were used for colouring the HYGIN network in Cytoscape. These terms were curated by and courtesy of the Vizeacoumar research group. The common name for the human protein and its accession number accompany the GO Slim terms that have been curated by the Vizeacoumar research group. *DDR Pathways = "DNA Damage and Repair Pathway" and NA Metabolism = "Nucleic Acid Metabolism".

Supplementary Table B.4: List of genes that were identified as down-regulated in breast cancer at a 2-fold cut-off and with a Wilcoxon Signed Rank P-value less than 0.05. Mean expression is the average of the log(2) ratio (cancer gene expression/normal gene expression) for each gene. P-values are from Wilcoxon Signed Rank test (Python) after adjustment (Benjamini-Hochberg procedure), and log(10)(P-values) were used to generate the volcano plot in Figure 5.3.

Supplementary Table B.5: Genetic interactions in the breast cancer subnetwork. The table contains a list of the SL gene pairs that make up the breast cancer specific subnetwork where column GeneA represents the genes that are down-regulated in breast cancer, and column GeneB are their SL gene pairs.

Appendix C

Permission to Reprint

This section contains the permission to reprint three figures that are present in this thesis.

C.1 Figure 2.1

Attached is the permission from Nature Reviews Genetics to replicate Figure 1 from Welcker and Clurman [130].

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Jul 29, 2018

This Agreement between University of Saskatchewan -- Morgan Kirzinger ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4398000010979
License date	Jul 29, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation
Licensed Content Author	Markus Welcker, Bruce E. Clurman
Licensed Content Date	Feb 1, 2008
Licensed Content Volume	8
Licensed Content Issue	2
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Humanized yeast genetic interaction mapping predicts synthetic lethal interactions in breast cancer
Instructor name	Dr. Anthony Kusalik
Institution name	University of Saskatchewan

Expected presentation date	Aug 2018
Portions	Figure 1
Requestor Location	University of Saskatchewan 176 Thorvaldson Bldg 110 Science Place Saskatoon, SK SK S7N 5C9 Canada Attn: University of Saskatchewan
Billing Type	Invoice
Billing Address	University of Saskatchewan 176 Thorvaldson Bldg 110 Science Place Saskatoon, SK SK S7N 5C9 Canada Attn: University of Saskatchewan
Total	0.00 CAD

Terms and Conditions

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.0

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

C.2 Figure 2.2

Attached is the permission from Cell Press to replicate Figure 1 from Paul *et al.* [89].

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 27, 2018

This Agreement between University of Saskatchewan -- Morgan Kirzinger ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4417250186269
License date	Aug 27, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Molecular Medicine
Licensed Content Title	Building high-resolution synthetic lethal networks: a 'Google map' of the cancer cell
Licensed Content Author	James M. Paul,Shaina D. Templeton,Akanksha Baharani,Andrew Freywald,Franco J. Vizeacoumar
Licensed Content Date	Dec 1, 2014
Licensed Content Volume	20
Licensed Content Issue	12
Licensed Content Pages	12
Start Page	704
End Page	715
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Humanized yeast genetic interaction mapping predicts synthetic lethal interactions in breast cancer
Publisher of new work	University of Saskatchewan
Author of new work	Dr. Anthony Kusalik
Expected completion date	Aug 2018
Estimated size (number of pages)	1
Requestor Location	University of Saskatchewan 176 Thorvaldson Bldg 110 Science Place Saskatoon, SK SK S7N 5C9

	Canada Attn: University of Saskatchewan
Publisher Tax ID	GB 494 6272 12
Total	0.00 CAD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the

formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.
Additional Terms & Conditions applicable to each Creative Commons user license:
CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.
CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.
Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.9

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

C.3 Figure 2.3

Attached is the permission from European Journal of Cancer to replicate Figure 1 from Ferrari *et al.* [34].

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 27, 2018

This Agreement between University of Saskatchewan -- Morgan Kirzinger ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4417241407360
License date	Aug 27, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	European Journal of Cancer
Licensed Content Title	A lethal combination for cancer cells: Synthetic lethality screenings for drug discovery
Licensed Content Author	Elisa Ferrari,Chiara Lucca,Marco Foiani
Licensed Content Date	Nov 1, 2010
Licensed Content Volume	46
Licensed Content Issue	16
Licensed Content Pages	7
Start Page	2889
End Page	2895
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Humanized yeast genetic interaction mapping predicts synthetic lethal interactions in breast cancer
Publisher of new work	University of Saskatchewan
Author of new work	Dr. Anthony Kusalik
Expected completion date	Aug 2018
Estimated size (number of pages)	1
Requestor Location	University of Saskatchewan 176 Thorvaldson Bldg 110 Science Place Saskatoon, SK SK S7N 5C9 Canada Attn: University of Saskatchewan
Publisher Tax ID	GB 494 6272 12

Total 0.00 CAD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all

claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of

articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.
Terms & Conditions applicable to all Open Access articles published with Elsevier:
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.
The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.
Additional Terms & Conditions applicable to each Creative Commons user license:
CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.
CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.
Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.
Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.9

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.